INTERNATIONAL JOURNAL OF TRANSGENDER HEALTH
2024, VOL. 25, NO. 3, 407–418
https://doi.org/10.1080/26895269.2023.2218357

Taylor & Francis
Taylor & Francis Group

REVIEW ARTICLE

OPEN ACCESS Check for updates

Randomized-controlled trials are methodologically inappropriate in adolescent transgender healthcare

Florence Ashley[a,*], Diana M. Tordoff[b], Johanna Olson-Kennedy[c,d] and Arjee J. Restar[e,f]

[a]Faculty of Law & Joint Centre for Bioethics, University of Toronto, Toronto, ON, Canada; [b]School of Medicine, Stanford University, Stanford, CA, USA; [c]The Center for Transyouth Health and Development, Division of Adolescent Medicine, Children's Hospital Los Angeles, Los Angeles, CA, USA; [d]Keck School of Medicine, University of Southern California, Los Angeles, CA, USA; [e]Department of Epidemiology, School of Public Health, University of Washington, Seattle, WA, USA; [f]Research Education Institute for Diverse Scholars (REIDS), School of Public Health, Yale University, New Haven, CT, USA

ABSTRACT

Background: Despite multiple rigorous observational studies documenting the association between positive mental health outcomes and access to puberty blockers, hormone therapy, and transition-related surgeries among adolescents, some jurisdictions have banned or are attempting to ban gender-affirming medical interventions for minors due to an absence of randomized-controlled trials (RCTs) proving their mental health benefits.

Methods: This article critically reviews whether RCTs are methodologically appropriate for studying the association between adolescent gender-affirming care and mental health outcomes.

Results: The scientific value of RCTs is severely impeded when studying the impact of gender-affirming care on the mental health of trans adolescent. Gender-affirming interventions have physiologically evident effects and are highly desired by participants, giving rise to concerns over adherence, drop-out, response bias, and generalizability. Complementary and well-designed observational studies can instead be used to ground reliable recommendations for clinical practice and policymaking in adolescent trans healthcare, without the need for RCTs.

Conclusion: The lack of RCTs on the mental health impacts of gender-affirming care for trans adolescents does not entail that gender-affirming interventions are based on insufficient evidence. Given the methodological limitations of RCTs, complementary and well-designed observational studies offer more reliable scientific evidence than RCTs and should be considered of sufficient quality to guide clinical practice and policymaking.

KEYWORDS

evidence-based medicine; gender-affirming care; randomized-controlled trials; research methods; transgender

Evidence of mental health benefits from puberty blockers, hormone therapy, and surgeries for transgender adolescents has been criticized for being based on 'low-quality' observational evidence (NICE, 2020a, 2020b). Opponents of gender-affirming care have argued that the evidence base is insufficient to justify the practices, leading some institutions and jurisdictions to restrict or criminalize access to gender-affirming care among minors (Caputo, 2022; Florida Medicaid, 2022). The assertion that gender-affirming care for transgender adolescents is based on low-quality evidence is predicated on the absence of randomized-controlled trials (RCTs) proving the mental health benefits of gender-affirming interventions. RCTs are considered the gold-standard of evidence-based medicine (Hariton & Locascio, 2018). For example, under the Grading of Recommendations, Assessment, Development and Evaluations (GRADE) framework for evaluating scientific evidence, only RCTs provide high-quality evidence by default, whereas observational studies are typically ranked as low-quality evidence (Guyatt et al., 2008). However, extensive scholarly literature has suggested that RCTs are not always a methodologically preferable or appropriate study design (Black, 1996; Bondemark & Ruf, 2015; Brewin & Bradley, 1989; Deaton & Cartwright, 2018; Feinstein & Horwitz, 1997; Frieden, 2017; Ginsburg & Smith, 2016; Grossman & Mackenzie, 2005; Kennedy-Martin et al., 2015; Marshall & Marshall, 2007; Mykhalovskiy

CONTACT Arjee J. Restar restar@uw.edu Department of Epidemiology, School of Public Health, University of Washington, Seattle, WA, USA.
*Current affiliation: Faculty of Law & John Dossetor Health Ethics Centre, University of Alberta, Edmonton, AB, Canada

& Weir, 2004; Sanson-Fisher et al., 2007; Simon, 2001). Are RCTs necessary for clinical decision-making and policymaking in adolescent transgender healthcare?

This article argues that RCTs are methodologically inappropriate for evaluating the impact of gender-affirming care on mental health outcomes among adolescents. The article is divided into three sections that (1) survey why and when RCTs are valued as a study design, (2) explain why RCTs are insufficiently methodologically robust for studying the impact of gender-affirming care on mental health among trans adolescents, and (3) discuss how complementary and well-crafted observational studies are an appropriate alternative basis for clinical practice and policymaking. Our arguments supplement ethical concerns regarding the lack of equipoise for RCTs in adolescent trans healthcare given the numerous studies associating adolescent gender-affirming interventions with improved mental health (Achille et al., 2020; Allen et al., 2019; Costa et al., 2015; Deutsch et al., 2016; de Vries et al., 2014, 2011; Foster Skewis et al., 2021; Freedman, 1987; Grannis et al., 2021; Green et al., 2022; Kuper et al., 2020; Sorbara et al., 2020; Tordoff, Wanta, et al., 2022; Tan et al., 2022; Turban et al., 2020, 2022).

Why and when RCTs are valued

RCTs involve randomizing participants between two (or more) groups, one of which receives the proposed intervention and one of which does not. RCTs have emerged as a gold standard for medical interventions due to their ability to infer causality, ensure exchangeability, and minimize bias from unmeasured confounders. Confounding occurs when the effect of the exposure or intervention (e.g. receiving hormone therapy) is mixed with the effect of another variable, leading to biased associations with the outcome of interest (e.g. reported quality of life). Because RCTs involve masking and randomization, they are designed to distribute confounders equally between the intervention and control groups, thereby minimizing bias due to unmeasured confounders (Sibbald & Roland, 1998). Observational studies are more vulnerable to unmeasured confounding because patients know whether they have received the intervention, typically have a say in whether they do so, and are not always compared to a control group (although many observational study designs do include comparison groups).

While RCTs are resilient to the influence of unmeasured confounding, they are vulnerable to other systematic errors resulting from impediments to participants adhering to the study protocol. Non-adherence includes inconsistent administration of the intervention, inconsistent data collection, and access to the intervention by participants in the control group. To address these problems, RCTs typically analyze participant data based on the groups participants were assigned to, regardless of any non-adherence. This approach, known as intention-to-treat analysis, presumes that non-adherence is an integral part of clinical reality and includes it to offer a more accurate picture of the intervention's effect in a clinical setting (Detry & Lewis, 2014; Gupta, 2011). However, intention-to-treat does not address all forms of systematic bias (Bell et al., 2013; Detry & Lewis, 2014; White et al., 2011). It does not address bias associated with missing outcome data due to participants withdrawing from the study, who may differ from participants who remained in the study. Nor does it address the bias that arises from non-adherence due to participants becoming aware of the group they were assigned to, which can lead studies to significantly underestimate the effect of an intervention.

Systematic bias from non-adherence and missing data is a recognized problem in studies that compare a new and untested intervention to an intervention believed to be effective, as patients assigned to the untested intervention may be more inclined to withdraw from a study or not adhere to the treatment protocol (Detry & Lewis, 2014; Piaggio et al., 2006). The risk of bias is minimized when participants and researchers remain unaware of who is receiving the intervention and who is part of the control group (Lewis & Warlow, 2004). Masking trial assignments can reduce the risk of bias in assessing outcomes of interest, such as mental health. However, it is impossible to maintain masking for interventions that have clear physiological or psychoactive effects that reveal the group to which participants were randomized.

Well-designed RCTs are powerful scientific tools. However, their value varies depending on the population being studied, the intervention and outcomes of interest, and the context of the study. RCTs trade one type of bias for others (Sullivan, 2011). While minimal unmasking and non-adherence are not fatal to the validity of RCTs, a significant degree of either limits their scientific utility. Besides impediments to internal validity, RCTs also suffer from threats to external validity. Findings from RCTs are not always generalizable to the population of interest because those who choose to participate in the study may not be representative of the population to whom the intervention would be applied in a clinical setting, limiting their usefulness in guiding clinical practices and policymaking (Hariton & Locascio, 2018; Kennedy-Martin et al., 2015).

Absent proper randomization, masking, adherence, and generalizability, RCTs fail to provide higher scientific evidence relative to well-conducted observational studies. In some areas of scientific research, RCTs are of limited value and are considered inappropriate (Ginsburg & Smith, 2016; Marshall & Marshall, 2007; Sullivan, 2011). As suggested by one scholar, "it would be a mistake to label the RCT as a gold standard for all research" (Simon, 2001).

The limitations of RCTs in adolescent transgender healthcare

RCTs are ill-suited to studying the effects of gender-affirming interventions on the psychological well-being and quality of life of transgender adolescents. Adequate masking, adherence, and generalizability are severely impeded when studying gender-affirming care for trans adolescents, limiting the scientific value of RCTs.

Gender-affirming interventions have physiologically evident effects, making it impossible to mask RCTs. The purpose of puberty blockers, hormone therapy, and transition-related surgeries is to inhibit or produce visible bodily changes. In an RCT, adolescents who are on puberty blockers would notice that their endogenous pubertal development had stopped, whereas those not on puberty blockers will notice that they had not. Similarly, adolescents given hormone therapy would notice bodily changes from taking estrogen or testosterone, whereas adolescents in the control arm would notice no such changes. Hormonal suppression is achieved around four weeks after treatment is initiated, but it may take multiple months before participants notice that pubertal development has ceased. The onset of visible effects from hormone therapy varies from person to person (Hembree et al., 2017). The first changes typically appear between one and six months of initiation, whereas other desired changes may not begin for up to a year. Although it may take some time before participants can ascertain the group, they were allocated to due to the delayed effect of puberty blockers and the progressive effect of hormone therapy, large-scale unmasking is inevitable. Because physiological changes are the primary purpose of gender-affirming care, meaningful effects on psychological well-being and quality of life are not expected until unmasking occurs. As such, while RCTs can be utilized to examine the effects of gender-affirming care on physiological changes, using RCTs to measure the effect of gender-affirming care on psychological well-being and quality of life is inappropriate.

Unmasking an RCT of gender-affirming care would lead to noncompliance, cross-over, and response bias in the control arm of the study. Adolescents who pursue gender-affirming care are typically insistent and persistent in seeking the interventions. They are not ambivalent as to whether they are assigned to the intervention or control arm of the study. Upon realizing that they are in the control arm due to physiological effects or lack thereof, a large proportion of the study participants would likely withdraw from the study or pursue alternative sources of gender-affirming interventions. This is especially true for puberty blockers given the undesired and irreversible nature of endogenous puberty for participants. Withdrawing from the study and non-adherence to the study protocol is most likely among adolescents who have alternative means of securing gender-affirming care and who experience more severe bodily gender dysphoria, raising grave concerns of systematic bias. Gender-affirming interventions can be obtained from parents, peers, illicit or unauthorized sources, other providers within

or outside the jurisdiction, and through medication-sharing with participants from the active arm of the study (Carlile et al., 2021; Horton, 2021, 2022; Rotondi et al., 2013). Some of these options are associated with elevated safety risks, giving rise to additional ethical concerns about the use of RCTs (Branstetter, 2016; Edenfield et al., 2019). Intentional withdrawal with the goal of forcing the study to end is also possible. Resentment toward researchers for not allowing all participants to receive gender-affirming interventions may also increase the risk of response bias compared to observational studies, and the experimental design may motivate adolescents to engage in self-harm or suicidal behavior to influence the study results, aggravating scientific and ethical concerns. Prior to unmasking, responses may also be biased due to the emotional stress associated with participants not knowing which arm of the study they were assigned to, leading the study to underestimate the mental health benefits of gender-affirming care. Given that withdrawal rates could be high enough for studies to be terminated before they are concluded, RCTs may prove impossible to conduct altogether. In one study on central precocious puberty, for example, parents of all participants assigned to the control group, which did not receive puberty blockers, withdrew from the study to pursue puberty blockers from another source, forcing the authors to change their study design (Mul et al., 2007). The likelihood of withdrawal, non-adherence, and response bias in the context of gender-affirming care for trans adolescents undermines RCTs' ability to detect true associations and avoid biased associations between the intervention and the outcomes.

Multiple baseline designs are occasionally used to address some of the ethical and methodological concerns associated with RCTs. Multiple baseline designs are similar to RCTs but randomize participants to two or more groups that initiate the intervention at different points in time instead of randomizing them to synchronized intervention and control arms (Hawkins et al., 2007; Rhoda et al., 2011). Multiple baseline designs are, however, not a viable alternative to RCTs due to the delayed and progressive effects of puberty blockers and hormone therapy. If the initiation interval between groups is short, the study would not reveal associations between the intervention and the outcomes. If the initiation interval between groups is long, the risk of systematic bias specific to withdrawal and non-adherence would remain. The risk of response bias remains regardless of the length of the initiation interval, given the clear physiological effects of interventions.

RCTs of gender-affirming interventions also give rise to problems of generalizability. Reluctance to participate is a known impediment to the scientific value of RCTs (Black, 1996; Detry & Lewis, 2014). We have reasons to believe that adolescents who consent or assent to participate in an RCT are unrepresentative of the clinical population. Adolescents who have alternative means of accessing gender-affirming care are unlikely to accept the risk of being randomized to the control group, given their settled desire for the intervention (Carlile et al., 2021; Horton, 2021, 2022). Adolescents who are white, socioeconomically privileged, live in areas with more gender-affirming care providers, and have strong parental support are less likely to participate (Everhart et al., 2022; Lett et al., 2022; Tordoff, Sequeira, et al., 2022). Adolescents who are desperate and/or have no other way of effectively pursuing gender-affirming interventions are more likely to participate in an RCT. Participation would also be impacted by the severity of the adolescent's bodily gender dysphoria, as those who experience greater distress toward their body are more likely to seek out guaranteed avenues of obtaining puberty blockers, hormone therapy, or surgeries and, if unsuccessful, may be more inclined to participate in the study out of desperation. Parental attitudes are also likely to impact parents' willingness to let their child participate, with more affirming parents being less willing to risk their child being assigned to the control arm, and parental acceptance has a known influence on the mental health of trans adolescents (Bauer et al., 2015; Jin et al., 2020; Simons et al., 2013). These considerations apply to both initial and ongoing participation, leading to systematic bias due to differential withdrawal after unmasking occurs.

Numerous studies associate gender-affirming care with improved mental health among adolescents (Achille et al., 2020; Allen et al., 2019; Costa et al., 2015; Deutsch et al., 2016; de Vries et al.,

2014, 2011; Foster Skewis et al., 2021; Freedman, 1987; Grannis et al., 2021; Green et al., 2022; Kuper et al., 2020; Sorbara et al., 2020; Tordoff, Wanta, et al., 2022; Turban et al., 2020, 2022). Researchers and clinicians who are convinced of the effectiveness of gender-affirming care, many of whom are leading providers in the field, are also unlikely to accept involvement with an RCT due to ethical concerns, further limiting and biasing the recruitment pool. These considerations suggest that RCTs of gender-affirming care would not be generalizable to the overall clinical population because their sample is not representative, unlike observational studies that may recruit any adolescent who seeks gender-affirming care.

The generalizability of results may also be undermined because of the homogeneity of interventions in explanatory RCTs, which aim at detecting causal relationships by emphasizing internal validity (Charlton, 1994; Patsopoulos, 2011; Roland & Torgerson, 1998). Each arm of an explanatory RCT typically offers identical or nearly identical interventions and settings since individualized interventions may violate exchangeability and re-introduce the risk of unmeasured confounders (Charlton, 1994; Roland & Torgerson, 1998). Standardized gender-affirming interventions are inconsistent with clinical practice, where gender-affirming interventions are individualized to the patient's embodiment goals and endocrine response, and may be offered for exploratory purposes. Some trans people prefer a low dose of testosterone instead of a typical dose to slow and/or adjust the bodily changes they incur (Bass et al., 2018). Hormone therapy dosages are routinely adjusted based on blood serum levels, biomarkers, and experienced side effects. Patients may try different formulations and medications before finding one that is suitable. Puberty blockers can be offered to adolescents who are uncertain of whether they want to initiate hormone therapy. Some adolescents initiate hormone therapy knowing that they only want to remain on hormones temporarily. Yet others may initiate hormone therapy because they are unsure of whether they will enjoy the effects and want to test them out (Ashley, 2019). These subgroups of patients are more likely to withdraw before the conclusion of the study and may be much less likely to participate in RCTs that adopt a homogeneous approach to gender-affirming care.

Due to the unique social and clinical context of adolescent gender-affirming care, RCTs examining the mental health outcomes of puberty blockers, hormone therapy, and transition-related surgeries cannot ensure adequate masking, adherence, and generalizability. Systematic bias is highly probable and would have a large predicted influence on results. Conducting RCTs is, therefore, methodologically inappropriate for studying the effects of gender-affirming care on the psychological well-being of trans adolescents.

The value of complementary observational studies

Complementary and well-designed observational studies are preferable to RCTs in adolescent trans healthcare given the above-described limitations. Although RCTs provide the most stringent criteria for inferring causality, the effectiveness of an intervention can be established through an accumulation of evidence from well-designed observational studies (Hernán, 2018). Moreover, observational studies are less vulnerable to the forms of systematic bias described above. By offering gender-affirming interventions to all participants who desire them, observational studies can reduce the risks associated with withdrawal and non-adherence. Because observational studies do not intervene on the clinical environment and merely observe it, they can more closely ensure generalizability to the clinical population. While observational studies are vulnerable to unmeasured confounding, methodological and statistical tools can be used to ascertain and limit the risk of unmeasured confounding and control for measured confounders. Complementary and well-designed observational studies can ground reliable recommendations for clinical practice and policymaking in adolescent trans healthcare, without the need for RCTs.

Many disciplines and areas of research rely on observational studies because RCTs are considered impracticable or unethical. This is especially common when studying the mental health outcomes of physiologically evident interventions due to the impossibility of masking, and when studying the outcomes of highly desired interventions due to the risks of de-randomization

(Brewin & Bradley, 1989). Psychological and psychosocial interventions are most commonly studied using observational methodologies, and many research questions remain unstudied with RCTs (Andrews, 1999; Stirman et al., 2003). Clinical recommendations relating to the mental health impacts of abortion, for instance, are based on observational studies rather than RCTs since abortion is a highly-desired intervention and is impossible to mask (Academy of Medical Royal Colleges, 2011; see also Ashley, 2022). The mental health outcomes of surgeries are typically not studied with RCTs due to the difficulty of masking, ethical concerns over sham or 'placebo' surgeries, and the belief that observational studies are sufficiently reliable to guide clinical practice (Baum, 1999; London & Kadane, 2002; Macklin, 1999; Reeves, 1999). Authors have explicitly recommended against RCTs in certain fields that involve psychosocial interventions (Ginsburg & Smith, 2016; Howick, 2020; Marshall & Marshall, 2007; Sullivan, 2011). The value of RCTs has also been questioned in population health (Sanson-Fisher et al., 2007).

Clinical decisions and policymaking are often based on observational evidence instead of RCTs (Frieden, 2017; Hernán, 2021). In a study of 608 systematic reviews, less than 13.5% of reviewed interventions were backed by a RCT (Fleming et al., 2016; see also Howick et al., 2020). In pediatric medical guidelines, off-label drug use is frequently recommended despite an absence of RCTs (Meng et al., 2022). Over half of the strong recommendations offered in World Health Organization guidelines are not based on RCTs (Alexander et al., 2014). According to Dr Thomas R. Frieden, former director of the *Centers for Disease Control and Prevention* (2017, p. 469): "For much, and perhaps most, of modern medical practice, RCT-based data are lacking, and no RCT is being planned or is likely to be completed to provide evidence for action." Routinely collected healthcare data is increasingly used as an alternative to RCTs in medicine (Hernán & Robins, 2016; Rogers et al., 2021). Government agencies recommend relying on observational evidence to determine the benefits and effects of interventions when RCTs would be impractical or infeasible (Norris et al., 2008). As a now-classic medical article points out, there are no RCTs demonstrating the effectiveness of parachutes (Smith & Pell, 2003). While RCTs are valuable, it is clear that the absence of RCTs has not been a general impediment to clinical decision-making or policymaking.

Concerns over the quality of evidence offered by observational studies are often overblown. Studies have shown that observational studies and RCTs tend to report comparable effect sizes when evaluating an intervention (Benson & Hartz, 2000; Ross, 2014). By contrast, a review of sampling practices has found that participants in most RCTs are "not broadly representative of patients treated in everyday clinical practice," raising serious concerns about the generalizability of RCTs that aim to inform clinical practice and policymaking (Kennedy-Martin et al., 2015). Taken together, these findings suggest that observational studies are a more reliable guide to clinical practice and policymaking than RCTs in research contexts that involve significant impediments to masking, randomization, and generalizability. Well-designed observational studies are often preferable to a fundamentally flawed RCT, despite being considered lower-quality evidence in evidence-based medicine.

While the risk of unmeasured confounding cannot be obviated, study design and statistical tools can be used to ascertain and mitigate the likelihood of systematic errors in observational studies. Observational studies can and should aim to control for temporal trends and measurable confounders. Quantitative bias analysis can be used to estimate the direction, magnitude, and uncertainty associated with systematic errors that can be introduced by unmeasured confounders, selection bias, and measurement errors (Lash et al., 2009, 2014). Statistical methods such as the E-value can be used to assess the minimum strength an unmeasured confounder would need in order to explain away a measured association (Haneuse et al., 2019; VanderWeele & Arah, 2011). For instance, a prospective cohort study by Tordoff and colleagues used E-values to establish that the reported outcome could only be explained away by "an unmeasured confounder that was associated with both [puberty blockers and hormone therapy] and the outcomes of interest by a risk ratio of 2-fold to

3-fold each, above and beyond the measured confounders" (Tordoff, Wanta, et al., 2022). An unforeseen association of this strength is unlikely, making the study's results more resistant to bias from unmeasured confounding and, thus, more reliable for clinical practice and policymaking.

The theoretical grounding, heterogeneity of studies, and complementary methodologies used in transgender health research can help rule out alternative explanations for observational results and enhance our ability to draw causal inferences. For instance, longitudinal studies showing improvements in mental health over time after receiving gender-affirming interventions make it less likely that cross-sectional studies showing better mental health among people who received gender-affirming interventions are explained by reverse causality or by individuals with a certain mental health profile being more likely to receive interventions due to the eligibility and readiness criteria used by clinicians. And in turn, contextual information suggests that temporal trends do not merely reflect patients returning to psychological baseline after a period of unusually low mood. Given the extreme waitlists at many youth gender identity clinics, which are already over 47 months in England, regression to the mean coinciding with initiating gender-affirming care after years of waiting would be a fantastic coincidence (Newport, 2021). Qualitative studies are also invaluable in interpreting observational studies, as participants can shed light on why certain associations were observed. Whereas quantitative studies tell us *what* happens, qualitative studies help tell us *why*.

Science is holistic and involves the testing and falsification of theories rather than isolated hypotheses (Duhem, 1976; Lakatos, 1976; Popper, 2005; Quine, 1951). Failing to reject a hypothesis after running the same study five times only provides marginal confidence over running the study once. Failing to refute five different hypotheses derived from the same theory, by contrast, meaningfully solidifies the entire theory. Without denying the value of replication, science progresses by testing theories across spaces of falsification. Each failed attempt at rejecting a hypothesis derived from the theory enhances scientific confidence in other hypotheses derived from the theory. Because gender-affirming care for transgender adolescents is grounded in the theory that affirming a person's sense of gender is beneficial, studies on adults as well as studies reporting benefits from social and legal gender affirmation bolster the scientific foundations of gender-affirming interventions for adolescents. While suboptimal, inferring from adult studies is common with pharmaceuticals. Each of these studies confirms the theory that gender affirmation is important to psychological well-being and quality of life.

Evaluations of an intervention's evidentiary foundations should reflect the feasibility and ethics of RCTs and take into consideration the scientific value of methodological heterogeneity in observational research. Frameworks of evidence-based medicine overvalue the differences between study types and undervalue differences within study types. While frameworks adapted to observational studies, such as the Newcastle-Ottawa scale, are more adapted to assessing the evidence base of gender-affirming interventions, they do not account for the scientific value of complementary study designs (Karalexi et al., 2020; Sterne et al., 2016). Standardized frameworks for evaluating evidence are prone to underestimate evidence from research programmes that use diverse methodologies and study designs compared to research programmes that repeatedly use the same methodology and study design.

Clear and convergent findings across diverse study designs establish a solid evidentiary foundation for clinical practice and policymaking in gender-affirming care, without having to resort to RCTs. Studies have included cross-sectional, prospective cohort, and qualitative studies. Studies have used varied sampling methods, control groups, and outcomes of interest at different points in time and stages of the transition process. These studies have consistently shown that neutral or beneficial mental health outcomes are associated with the receipt of gender-affirming care (Ashley, 2022). While long-term follow-up studies are needed to better understand the impact of gender-affirming interventions across the life course, existing data do not suggest that these risks exceed those posed by other routine

medical care. The evidence in favor of gender-affirming care is sufficiently strong and convergent to establish puberty blockers, hormone therapy, and transition-related surgeries as necessary care for adolescents who desire them.

Conclusion

The absence of RCTs studying the impact of gender-affirming care on the mental health and well-being of transgender adolescents does not imply that these interventions are insufficiently supported by evidence. Although RCTs are considered high-quality evidence because of their ability to control for unmeasured confounders, the impossibility of masking which participants receive gender-affirming interventions and the differential impact of unmasking on adherence, withdrawal, response bias, and generalizability compromises the value of RCTs for adolescent gender-affirming care. RCTs are methodologically inappropriate for studying the relationship between gender-affirming interventions and mental health. These methodological considerations compound the serious ethical concerns raised by RCTs in adolescent transgender healthcare. Given the limitations of RCTs, complementary and well-designed observational studies offer more reliable scientific evidence than RCTs and should be considered of sufficient quality to guide clinical practice and policymaking. Adolescent trans healthcare is on solid footing.

Conflicts of interest

The authors have no conflicts of interest relevant to this article to disclose.

Funding

Dr. Restar is supported by the Research Education Institute for Diverse Scholars (REIDS) Program at Yale University School of Public Health, funded by the National Institute of Mental Health (R25MH087217). This article does not represent the official view of the sponsor.

References

Academy of Medical Royal Colleges. (2011). *Induced abortion and mental health: A systematic review of the mental health outcomes of induced abortion, including their prevalence and associated factors*. https://perma.cc/N8ZP-L77C

Achille, C., Taggart, T., Eaton, N. R., Osipoff, J., Tafuri, K., Lane, A., & Wilson, T. A. (2020). Longitudinal impact of gender-affirming endocrine intervention on the mental health and well-being of transgender youths: Preliminary results. *International Journal of Pediatric Endocrinology*, *2020*(1), 8. https://doi.org/10.1186/s13633-020-00078-2

Alexander, P. E., Bero, L., Montori, V. M., Brito, J. P., Stoltzfus, R., Djulbegovic, B., Neumann, I., Rave, S., & Guyatt, G. (2014). World Health Organization recommendations are often strong based on low confidence in effect estimates. *Journal of Clinical Epidemiology*, *67*(6), 629–634. https://doi.org/10.1016/j.jclinepi.2013.09.020

Allen, L. R., Watson, L. B., Egan, A. M., & Moser, C. N. (2019). Well-being and suicidality among transgender youth after gender-affirming hormones. *Clinical Practice in Pediatric Psychology*, *7*(3), 302–311. https://doi.org/10.1037/cpp0000288

Andrews, G. (1999). Randomised controlled trials in psychiatry: Important but poorly accepted. *BMJ*, *319*(7209), 562–564. https://doi.org/10.1136/bmj.319.7209.562

Ashley, F. (2019). Thinking an ethics of gender exploration: Against delaying transition for transgender and gender creative youth. *Clinical Child Psychology and Psychiatry*, *24*(2), 223–236. https://doi.org/10.1177/1359104519836462

Ashley, F. (2022). Adolescent medical transition is ethical: An analogy with reproductive health. *Kennedy Institute of Ethics Journal*, *32*(2), 127–171. https://doi.org/10.1353/ken.2022.0010

Bass, M., Gonzalez, L. J., Colip, L., Sharon, N., & Conklin, J. (2018). Rethinking gender: The nonbinary approach. *American Journal of Health-System Pharmacy*, *75*(22), 1821–1823. https://doi.org/10.2146/ajhp180236

Bauer, G. R., Scheim, A. I., Pyne, J., Travers, R., & Hammond, R. (2015). Intervenable factors associated with suicide risk in transgender persons: A respondent driven sampling study in Ontario, Canada. *BMC Public Health*, *15*(1), art. 525. https://doi.org/10.1186/s12889-015-1867-2

Baum, M. (1999). Reflections on randomised controlled trials in surgery. *The Lancet*, *353*, S6–S8. https://doi.org/10.1016/S0140-6736(99)90220-9

Bell, M. L., Kenward, M. G., Fairclough, D. L., & Horton, N. J. (2013). Differential dropout and bias in randomised controlled trials: When it matters and when it may not. *BMJ*, *346*, e8668. https://doi.org/10.1136/bmj.e8668

Benson, K., & Hartz, A. J. (2000). A comparison of observational studies and randomized, controlled trials. *The New England Journal of Medicine*, *342*(25), 1878–1886. https://doi.org/10.1056/NEJM200006223422506

Black, N. (1996). Why we need observational studies to evaluate the effectiveness of health care. *BMJ*, *312*(7040), 1215–1218. https://doi.org/10.1136/bmj.312.7040.1215

Bondemark, L., & Ruf, S. (2015). Randomized controlled trial: The gold standard or an unobtainable fallacy?

European Journal of Orthodontics, *37*(5), 457–461. https://doi.org/10.1093/ejo/cjv046

Branstetter, G. (2016, August 31). Sketchy Pharmacies Are Selling Hormones to Transgender People. The Atlantic. https://www.theatlantic.com/health/archive/2016/08/diy-hormone-replacement-therapy/498044/

Brewin, C. R., & Bradley, C. (1989). Patient preferences and randomised clinical trials. *BMJ*, *299*(6694), 313–315. https://doi.org/10.1136/bmj.299.6694.313

Caputo, M. (2022, June 2). DeSantis moves to ban transition care for transgender youths, Medicaid recipients. NBC News. https://www.nbcnews.com/politics/politics-news/desantis-moves-ban-transition-care-transgender-youth-medicaid-recipien-rcna31736

Carlile, A., Butteriss, E., & Sansfaçon, A. P. (2021). "It's like my kid came back overnight": Experiences of trans and non-binary young people and their families seeking, finding and engaging with clinical care in England. *International Journal of Transgender Health*, *22*(4), 1–17. https://doi.org/10.1080/26895269.2020.1870188

Charlton, B. G. (1994). Understanding randomized controlled trials: Explanatory or pragmatic? *Family Practice*, *11*(3), 243–244. https://doi.org/10.1093/fampra/11.3.243

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*(11), 2206–2214. https://doi.org/10.1111/jsm.13034

de Vries, A. L. C., McGuire, J. K., Steensma, T. D., Wagenaar, E. C. F., Doreleijers, T. A. H., & Cohen-Kettenis, P. T. (2014). Young adult psychological outcome after puberty suppression and gender reassignment. *Pediatrics*, *134*(4), 696–704. https://doi.org/10.1542/peds.2013-2958

de Vries, A. L. C., Steensma, T. D., Doreleijers, T. A. H., & Cohen-Kettenis, P. T. (2011). Puberty suppression in adolescents with gender identity disorder: A prospective follow-up study. *The Journal of Sexual Medicine*, *8*(8), 2276–2283. https://doi.org/10.1111/j.1743-6109.2010.01943.x

Deaton, A., & Cartwright, N. (2018). Understanding and misunderstanding randomized controlled trials. *Social Science & Medicine (1982)*, *210*, 2–21. https://doi.org/10.1016/j.socscimed.2017.12.005

Detry, M. A., & Lewis, R. J. (2014). The intention-to-treat principle: How to assess the true effect of choosing a medical treatment. *JAMA*, *312*(1), 85–86. https://doi.org/10.1001/jama.2014.7523

Deutsch, M. B., Radix, A., & Reisner, S. (2016). What's in a Guideline? Developing collaborative and sound research designs that substantiate best practice recommendations for transgender health care. *The AMA Journal of Ethic*, *18*(11), 1098–1106. https://doi.org/10.1001/journalofethics.2016.18.11.stas1-1611

Duhem, P. (1976). Physical theory and experiment. In S. G. Harding (Ed.), *Can theories be refuted?* (pp. 1–40). Springer Netherlands. https://doi.org/10.1007/978-94-010-1863-0_1

Edenfield, A. C., Holmes, S., & Colton, J. S. (2019). Queering tactical technical communication: DIY HRT. *Technical Communication Quarterly*, *28*(3), 177–191. https://doi.org/10.1080/10572252.2019.1607906

Everhart, A. R., Ferguson, L., & Wilson, J. P. (2022). Construction and validation of a spatial database of providers of transgender hormone therapy in the US. *Social Science & Medicine (1982)*, *303*, 115014. https://doi.org/10.1016/j.socscimed.2022.115014

Feinstein, A. R., & Horwitz, R. I. (1997). Problems in the "Evidence" of "Evidence-based medicine. *The American Journal of Medicine*, *103*(6), 529–535. https://doi.org/10.1016/S0002-9343(97)00244-1

Fleming, P. S., Koletsi, D., Ioannidis, J. P. A., & Pandis, N. (2016). High quality of the evidence for medical and other health-related interventions was uncommon in Cochrane systematic reviews. *Journal of Clinical Epidemiology*, *78*, 34–42. https://doi.org/10.1016/j.jclinepi.2016.03.012

Medicaid, F. (2022). Generally accepted professional medical standards determination on the treatment of gender dysphoria. Florida Agency for Health Care Administration. https://www.ahca.myflorida.com/letkidsbekids/docs/AHCA_GAPMS_June_2022_Report.pdf

Foster Skewis, L., Bretherton, I., Leemaqz, S. Y., Zajac, J. D., & Cheung, A. S. (2021). Short-term effects of gender-affirming hormone therapy on dysphoria and quality of life in transgender individuals: A prospective controlled study. *Frontiers in Endocrinology*, *12*, 717766. https://doi.org/10.3389/fendo.2021.717766

Freedman, B. (1987). Equipoise and the ethics of clinical research. *The New England Journal of Medicine*, *317*(3), 141–145. https://doi.org/10.1056/NEJM198707163170304

Frieden, T. R. (2017). Evidence for health decision making—Beyond randomized, controlled trials. *The New England Journal of Medicine*, *377*(5), 465–475. https://doi.org/10.1056/NEJMra1614394

Ginsburg, A., Smith, M. S. (2016). Do randomized controlled trials meet the "Gold Standard"? A study of the usefulness of RCTs in the what works clearinghouse. American Enterprise Institute. https://www.carnegiefoundation.org/wp-content/uploads/2016/03/Do-randomized-controlled-trials-meet-the-gold-standard.pdf

Grannis, C., Leibowitz, S., Gahn, S., Nahata, L., Morningstar, M., Mattson, W. I., Chen, D., Strang, J. F., & Nelson, E. E. (2021). Testosterone treatment, internalizing symptoms, and body image dissatisfaction in transgender boys. *Psychoneuroendocrinology*, *132*, 105358. https://doi.org/10.1016/j.psyneuen.2021.105358

Green, A. E., DeChants, J. P., Price, M. N., & Davis, C. K. (2022). Association of gender-affirming hormone therapy with depression, thoughts of suicide, and attempted suicide among transgender and nonbinary youth. *Journal of Adolescent Health*, *70*(4), 643–649. https://doi.org/10.1016/j.jadohealth.2021.10.036

Grossman, J., & Mackenzie, F. J. (2005). The randomized controlled trial: Gold standard, or merely standard?

Perspectives in Biology and Medicine, *48*(4), 516–534. https://doi.org/10.1353/pbm.2005.0092

Gupta, S. (2011). Intention-to-treat concept: A review. *Perspectives in Clinical Research*, *2*(3), 109. https://doi.org/10.4103/2229-3485.83221

Guyatt, G. H., Oxman, A. D., Vist, G. E., Kunz, R., Falck-Ytter, Y., Alonso-Coello, P., & Schünemann, H. J, GRADE Working Group (2008). GRADE: An emerging consensus on rating quality of evidence and strength of recommendations. *BMJ*, *336*(7650), 924–926. https://doi.org/10.1136/bmj.39489.470347.AD

Haneuse, S., VanderWeele, T. J., & Arterburn, D. (2019). Using the E-value to assess the potential effect of unmeasured confounding in observational studies. *JAMA*, *321*(6), 602–603. https://doi.org/10.1001/jama.2018.21554

Hariton, E., & Locascio, J. J. (2018). Randomised controlled trials—The gold standard for effectiveness research. *BJOG: An International Journal of Obstetrics and Gynaecology*, *125*(13), 1716–1716. https://doi.org/10.1111/1471-0528.15199

Hawkins, N. G., Sanson-Fisher, R. W., Shakeshaft, A., D'Este, C., & Green, L. W. (2007). The multiple baseline design for evaluating population-based research. *American Journal of Preventive Medicine*, *33*(2), 162–168. https://doi.org/10.1016/j.amepre.2007.03.020

Hembree, W. C., Cohen-Kettenis, P. T., Gooren, L., Hannema, S. E., Meyer, W. J., Murad, M. H., Rosenthal, S. M., Safer, J. D., Tangpricha, V., & T'Sjoen, G. G. (2017). Endocrine treatment of gender-dysphoric/gender-incongruent persons: An Endocrine Society* Clinical Practice Guideline. *The Journal of Clinical Endocrinology and Metabolism*, *102*(11), 3869–3903. https://doi.org/10.1210/jc.2017-01658

Hernán, M. A. (2018). The C-Word: Scientific euphemisms do not improve causal inference from observational data. *American Journal of Public Health*, *108*(5), 616–619. https://doi.org/10.2105/AJPH.2018.304337

Hernán, M. A. (2021). Methods of public health research—Strengthening causal inference from observational data. *The New England Journal of Medicine*, *385*(15), 1345–1348. https://doi.org/10.1056/NEJMp2113319

Hernán, M. A., & Robins, J. M. (2016). Using big data to emulate a target trial when a randomized trial is not available: Table 1. *American Journal of Epidemiology*, *183*(8), 758–764. https://doi.org/10.1093/aje/kwv254

Horton, C. (2021). "It felt like they were trying to destabilise us": Parent assessment in UK children's gender services. *International Journal of Transgender Health*, *24*(1), 70–85. https://doi.org/10.1080/26895269.2021.2004569

Horton, C. (2022). Experiences of puberty and puberty blockers: Insights from trans children, trans adolescents, and their parents. *Journal of Adolescent Research*. Advance online publication. https://doi.org/10.1177/07435584221100591

Howick, J. (2020, September 2). Only one in ten medical treatments are backed by high-quality evidence. The Conversation. https://theconversation.com/only-one-in-ten-medical-treatments-are-backed-by-high-quality-evidence-145224

Howick, J., Koletsi, D., Pandis, N., Fleming, P. S., Loef, M., Walach, H., Schmidt, S., & Ioannidis, J. P. A. (2020). The quality of evidence for medical interventions does not improve or worsen: A metaepidemiological study of Cochrane reviews. *Journal of Clinical Epidemiology*, *126*, 154–159. https://doi.org/10.1016/j.jclinepi.2020.08.005

Jin, H., Restar, A., Goedel, W. C., Ogunbajo, A., Biello, K., Operario, D., Kuhns, L., Reisner, S. L., Garofalo, R., & Mimiaga, M. J. (2020). Maternal support is protective against suicidal ideation among a diverse cohort of young transgender women. *LGBT Health*, *7*(7), 349–357. https://doi.org/10.1089/lgbt.2020.0219

Karalexi, M. A., Georgakis, M. K., Dimitriou, N. G., Vichos, T., Katsimpris, A., Petridou, E. T., & Papadopoulos, F. C. (2020). Gender-affirming hormone treatment and cognitive function in transgender young adults: A systematic review and meta-analysis. *Psychoneuroendocrinology*, *119*, 104721. https://doi.org/10.1016/j.psyneuen.2020.104721

Kennedy-Martin, T., Curtis, S., Faries, D., Robinson, S., & Johnston, J. (2015). A literature review on the representativeness of randomized controlled trial samples and implications for the external validity of trial results. *Trials*, *16*(1), art. 495. https://doi.org/10.1186/s13063-015-1023-4

Kuper, L. E., Stewart, S., Preston, S., Lau, M., & Lopez, X. (2020). Body dissatisfaction and mental health outcomes of youth on gender-affirming hormone therapy. *Pediatrics*, *145*(4), e20193006. https://doi.org/10.1542/peds.2019-3006

Lakatos, I. (1976). Falsification and the methodology of scientific research programmes. In S. G. Harding (Ed.), *Can theories be refuted?* (pp. 205–259). Springer Netherlands. https://doi.org/10.1007/978-94-010-1863-0_14

Lash, T. L., Fox, M. P., & Fink, A. K. (2009). *Applying quantitative bias analysis to epidemiologic data*. Springer.

Lash, T. L., Fox, M. P., MacLehose, R. F., Maldonado, G., McCandless, L. C., & Greenland, S. (2014). Good practices for quantitative bias analysis. *International Journal of Epidemiology*, *43*(6), 1969–1985. https://doi.org/10.1093/ije/dyu149

Lett, E., Abrams, M. P., Gold, A., Fullerton, F.-A., & Everhart, A. (2022). Ethnoracial inequities in access to gender-affirming mental health care and psychological distress among transgender adults. *Social Psychiatry and Psychiatric Epidemiology*, *57*(5), 963–971. https://doi.org/10.1007/s00127-022-02246-6

Lewis, S. C., & Warlow, C. P. (2004). How to spot bias and other potential problems in randomised controlled trials. *Journal of Neurology, Neurosurgery, and Psychiatry*, *75*(2), 181–187. https://doi.org/10.1136/jnnp.2003.025833

London, A. J., & Kadane, J. B. (2002). Placebos that harm: Sham surgery controls in clinical trials. *Statistical Methods in Medical Research*, *11*(5), 413–427. https://doi.org/10.1191/0962280202sm300ra

Macklin, R. (1999). The ethical problems with sham surgery in clinical research. *The New England Journal of Medicine*, *341*(13), 992–996. https://doi.org/10.1056/NEJM199909233411312

Marshall, W. L., & Marshall, L. E. (2007). The utility of the random controlled trial for evaluating sexual offender treatment: The gold standard or an inappropriate strategy? *Sexual Abuse: A Journal of Research and Treatment*, *19*(2), 175–191. https://doi.org/10.1177/107906320701900207

Meng, M., Zhou, Q., Lei, W., Tian, M., Wang, P., Liu, Y., Sun, Y., Chen, Y., & Li, Q. (2022). Recommendations on off-label drug use in pediatric guidelines. *Frontiers in Pharmacology*, *13*, 892574. https://doi.org/10.3389/fphar.2022.892574

Mul, D., Versluis-den Bieman, H., Slijper, F., Oostdijk, W., Waelkens, J., & Drop, S. (2007). Psychological assessments before and after treatment of early puberty in adopted children. *Acta Paediatrica*, *90*(9), 965–971. https://doi.org/10.1111/j.1651-2227.2001.tb01349.x

Mykhalovskiy, E., & Weir, L. (2004). The problem of evidence-based medicine: Directions for social science. *Social Science & Medicine (1982)*, *59*(5), 1059–1069. https://doi.org/10.1016/j.socscimed.2003.12.002

Newport, O. (2021, March 18). Current wait times for Tavistock and Portman Gender Identity Clinic. WhatDoTheyKnow. https://perma.cc/E54Q-JYML

NICE (2020a). Evidence review: Gender-affirming hormones for children and adolescents with gender dysphoria.

NICE (2020b). Evidence review: Gonadotrophin releasing hormone analogues for children and adolescents with gender dysphoria.

Norris, S., Atkins, D., Bruening, W., Fox, S., Johnson, E., Kane, R., Morton, S. C., Oremus, M., Ospina, M., Randhawa, G., Schoelles, K., Shekelle, P., & Viswanathan, M. (2008). Selecting observational studies for comparing medical interventions. In *Methods guide for effectiveness and comparative effectiveness reviews*. Agency for Healthcare Research and Quality. http://www.ncbi.nlm.nih.gov/books/NBK47093/

Patsopoulos, N. A. (2011). A pragmatic view on pragmatic trials. *Dialogues in Clinical Neuroscience*, *13*(2), 217–224. https://doi.org/10.31887/DCNS.2011.13.2/npatsopoulos

Piaggio, G., Elbourne, D. R., Altman, D. G., Pocock, S. J., & Evans, S. J. W, for the CONSORT Group. (2006). Reporting of noninferiority and equivalence randomized trials: An extension of the CONSORT statement. *JAMA*, *295*(10), 1152–1160. https://doi.org/10.1001/jama.295.10.1152

Popper, K. (2005). *The logic of scientific discovery*. Routledge. https://doi.org/10.4324/9780203994627

Quine, W. V. O. (1951). Two dogmas of empiricism. *The Philosophical Review*, *60*(1), 20–43. https://doi.org/10.2307/2181906

Reeves, B. (1999). Health-technology assessment in surgery. *The Lancet*, *353*, S3–S5. https://doi.org/10.1016/S0140-6736(99)90413-0

Rhoda, D. A., Murray, D. M., Andridge, R. R., Pennell, M. L., & Hade, E. M. (2011). Studies with staggered starts: Multiple baseline designs and group-randomized trials. *American Journal of Public Health*, *101*(11), 2164–2169. https://doi.org/10.2105/AJPH.2011.300264

Rogers, J. R., Lee, J., Zhou, Z., Cheung, Y. K., Hripcsak, G., & Weng, C. (2021). Contemporary use of real-world data for clinical trial conduct in the United States: A scoping review. *Journal of the American Medical Informatics Association: JAMIA*, *28*(1), 144–154. https://doi.org/10.1093/jamia/ocaa224

Roland, M., & Torgerson, D. J. (1998). Understanding controlled trials: What are pragmatic trials? *BMJ*, *316*(7127), 285–285. https://doi.org/10.1136/bmj.316.7127.285

Ross, J. S. (2014). Randomized clinical trials and observational studies are more often alike than unlike. *JAMA Internal Medicine*, *174*(10), 1557. https://doi.org/10.1001/jamainternmed.2014.3366

Rotondi, N. K., Bauer, G. R., Scanlon, K., Kaay, M., Travers, R., & Travers, A. (2013). Nonprescribed hormone use and self-performed surgeries: "Do-it-yourself" transitions in transgender communities in Ontario, Canada. *American Journal of Public Health*, *103*(10), 1830–1836. https://doi.org/10.2105/AJPH.2013.301348

Sanson-Fisher, R. W., Bonevski, B., Green, L. W., & D'Este, C. (2007). Limitations of the randomized controlled trial in evaluating population-based health interventions. *American Journal of Preventive Medicine*, *33*(2), 155–161. https://doi.org/10.1016/j.amepre.2007.04.007

Sibbald, B., & Roland, M. (1998). Understanding controlled trials: Why are randomised controlled trials important? *BMJ*, *316*(7126), 201–201. https://doi.org/10.1136/bmj.316.7126.201

Simon, S. D. (2001). Is the randomized clinical trial the gold standard of research? *Journal of Andrology*, *22*(6), 938–943. https://doi.org/10.1002/j.1939-4640.2001.tb03433.x

Simons, L., Schrager, S. M., Clark, L. F., Belzer, M., & Olson, J. (2013). Parental support and mental health among transgender adolescents. *Journal of Adolescent Health*, *53*(6), 791–793. https://doi.org/10.1016/j.jadohealth.2013.07.019

Smith, G. C. S., & Pell, J. P. (2003). Parachute use to prevent death and major trauma related to gravitational challenge: Systematic review of randomised controlled trials. *BMJ*, *327*(7429), 1459–1461. https://doi.org/10.1136/bmj.327.7429.1459

Sorbara, J. C., Chiniara, L. N., Thompson, S., & Palmert, M. R. (2020). Mental health and timing of gender-affirming care. *Pediatrics*, *146*(4), e20193600. https://doi.org/10.1542/peds.2019-3600

Sterne, J. A., Hernán, M. A., Reeves, B. C., Savović, J., Berkman, N. D., Viswanathan, M., Henry, D., Altman, D. G., Ansari, M. T., Boutron, I., Carpenter, J. R., Chan, A.-W., Churchill, R., Deeks, J. J., Hróbjartsson, A., Kirkham, J., Jüni, P., Loke, Y. K., Pigott, T. D., … Higgins, J. P. (2016). ROBINS-I: A tool for assessing risk of bias in non-randomised studies of interventions. *BMJ*, *355*, art. i4919.https://doi.org/10.1136/bmj.i4919

Stirman, S. W., DeRubeis, R. J., Crits-Christoph, P., & Brody, P. E. (2003). Are samples in randomized controlled trials of psychotherapy representative of community outpatients? A new methodology and initial findings. *Journal*

of Consulting and Clinical Psychology, *71*(6), 963–972. https://doi.org/10.1037/0022-006X.71.6.963

Sullivan, G. M. (2011). Getting off the "Gold Standard": Randomized controlled trials and education research. *Journal of Graduate Medical Education*, *3*(3), 285–289. https://doi.org/10.4300/JGME-D-11-00147.1

Tan, K. K. H., Byrne, J. L., Treharne, G. J., & Veale, J. F. (2022). Unmet need for gender-affirming care as a social determinant of mental health inequities for transgender youth in Aotearoa/New Zealand. *Journal of Public Health*. Advance online publication. https://doi.org/10.1093/pubmed/fdac131

Tordoff, D. M., Sequeira, G. M., Shook, A. G., Williams, F., Hayden, L., Kasenic, A., Inwards-Breland, D., & Ahrens, K. (2022). Factors associated with time to receiving gender-affirming hormones and puberty blockers at a pediatric clinic serving transgender and nonbinary youth. *Transgender Health*. Advance online publication. https://doi.org/10.1089/trgh.2021.0116

Tordoff, D. M., Wanta, J. W., Collin, A., Stepney, C., Inwards-Breland, D. J., & Ahrens, K. (2022). Mental health outcomes in transgender and nonbinary youths receiving gender-affirming care. *JAMA Network Open*, *5*(2), e220978. https://doi.org/10.1001/jamanetworkopen.2022.0978

Turban, J. L., King, D., Carswell, J. M., & Keuroghlian, A. S. (2020). Pubertal suppression for transgender youth and risk of suicidal ideation. *Pediatrics*, *145*(2), e20191725. https://doi.org/10.1542/peds.2019-1725

Turban, J. L., King, D., Kobe, J., Reisner, S. L., & Keuroghlian, A. S. (2022). Access to gender-affirming hormones during adolescence and mental health outcomes among transgender adults. *PloS One*, *17*(1), e0261039. https://doi.org/10.1371/journal.pone.0261039

VanderWeele, T. J., & Arah, O. A. (2011). Bias formulas for sensitivity analysis of unmeasured confounding for general outcomes, treatments, and confounders. *Epidemiology*, *22*(1), 42–52. https://doi.org/10.1097/EDE.0b013e3181f74493

White, I. R., Horton, N. J., Carpenter, J., & Pocock, S. J. (2011). Strategy for intention to treat analysis in randomised trials with missing outcome data. *BMJ*, *342*(feb07 1), d40. https://doi.org/10.1136/bmj.d40