

Critically Appraising the Cass Report: Methodological Flaws and Unsupported Claims

Chris Noone^{a,*}, Alex Southgate^c, Alex Ashman^d, Éile Quinn^b, David Comer^a, Duncan Shrewsbury^e, Florence Ashley^f, Jo Hartland^g, Joanna Pashdag^h, John Gilmoreⁱ, Natacha Kennedy^j, Thomas Woolley^k, Rachel Heath^l, Ryan Goulding^m, Victoria Simpsonⁿ, Ed Kiely^o, Sibéal Coll^p, Margaret White^q, D.M. Grijseels^r, Maxence Ouafik^s, Quinnehtukqut McLamore^t

^a*School of Psychology, University of Galway, University Road, H91 TK33, Galway, Ireland*

^b*School of Health Sciences, University of Galway, University Road, H91 TK33, Galway, Ireland*

^c*School of Physics & Astronomy, Cardiff University, CF10 3AT, Cardiff, United Kingdom*

^d*Thames Valley School of Surgery, United Kingdom*

^e*Department of Medical Education, Brighton & Sussex Medical School, BN1 9PX, Brighton, United Kingdom*

^f*University of Alberta Faculty of Law, AB T6G 2R3, Edmonton, Canada*

^g*Bristol Medical School, University of Bristol, 5 Tyndall Ave, BS8 1UD, Bristol, United Kingdom*

^h*Department of Psychology, Hawai'i Pacific University, 1 Aloha Tower Drive, HI 96813, Honolulu, United States*

ⁱ*School of Nursing, Midwifery and Health Systems, University College Dublin, 4 Stillorgan Rd, D04 C1P1, Belfield, Dublin, Ireland*

^j*Department of Educational Studies, Goldsmiths, 8 Lewisham Way, SE14 6NW, London, United Kingdom*

^k*School of Mathematics, Cardiff University, CF10 3AT, Cardiff, United Kingdom*

^l*School of Psychological Sciences, University of Newcastle, Wallace Street, NE24DR, Newcastle upon Tyne, Australia*

^m*School of Nursing and Midwifery, University College Cork, College Rd, T12 K8AF, Cork, Ireland*

ⁿ*Department of Psychology, Lancaster University, LA1 4YF, Bailrigg, Lancaster, United Kingdom*

^o*School of Geography, Queen Mary University of London, Mile End Rd, E1 4NS, London, United Kingdom*

^p*Trans Healthcare Action, Ireland*

^q*NHS Lothian, Waverley Gate, EH1 3EG, Edinburgh, Scotland*

^r*Max Planck Institute for Brain Research, Max-von-Laue-Straße 4, 60438, Frankfurt, Germany*

^s*General Practice Department, University of Liège, Quartier Hopital B23, Avenue Hippocrate 13, 4000, Liège, Belgium*

^t*Department of Psychological Sciences, University of Missouri at Columbia, McAlester Hall, 210, 320 S 6th St, MO 65201, Columbia, United States*

*Corresponding author: chris.noone@universityofgalway.ie

Abstract

The Cass Report aimed to provide recommendations for how services for gender diverse children and young people should be delivered in England. Our critical appraisal reveals significant methodological and conceptual flaws within the report and the research commissioned to inform the report, which included seven systematic reviews and both quantitative and qualitative primary research. Using the ROBIS tool, we identified a high risk of bias in each of the systematic reviews driven by unexplained protocol deviations, ambiguous eligibility criteria, inadequate study identification, and the failure to integrate consideration of these limitations into the conclusions derived from the evidence syntheses. We also identified potential sources of bias and unsubstantiated claims in the primary research that suggest a double standard in the quality of evidence produced for the Cass Report compared to quality appraisal in the systematic reviews. We discuss these issues in relation to how evidence regarding gender affirming care is framed, the wider political context, and the future for gender affirming care. To uphold evidence-based medicine, future gender-affirming care research must generate robust observational data, involve transgender communities, and prioritise patient-centred outcomes, ensuring validity, generalisability, and cultural relevance. The Cass Report's recommendations, given its methodological flaws and misrepresentation of evidence, warrant critical scrutiny to ensure ethical and effective support for gender-diverse youth.

Keywords: Cass Report, Gender Affirming Care, Gender Dysphoria, Gender Incongruence, Transgender

The Cass Report was commissioned by NHS England and NHS Improvement “to make recommendations on the services provided to children and young people who are exploring their gender identity or experiencing gender incongruence”¹. These recommendations were informed through seven systematic reviews, a survey of gender services for young people in Europe, two quantitative studies of health records, and a qualitative study on the experience of gender dysphoria (GD) among young people. In this paper, we show that this programme of research has significant methodological problems; that the Cass Report's application of evidence-based medicine (EBM) to Gender-Affirming Care (GAC) is flawed; and that the Review's understanding of transgender identities and experiences deploys a paternalistic lens that disregards the competence of transgender young people. We contend that together these issues undermine the validity of the Review's recommendations and that it has therefore failed to fulfil the commission request. We also make suggestions for how future research and systematic reviews on GAC should develop to ensure evidence-based quality improvement in the provision of GAC for transgender and gender diverse children and young people.

1. The Systematic Reviews

Seven systematic reviews were commissioned by the Cass Report and published in the BMJ Archives of Disease in Childhood²⁻⁸ with a single systematic review protocol registered in PROSPERO for all seven reviews⁹. To evaluate their risk of bias we applied the ROBIS tool¹⁰ as recommended by Cochrane^{11,12}. The ROBIS tool considers four domains of risk of bias: (1) study eligibility criteria, (2) identification and selection of studies, (3) data collection and study appraisal, and (4) synthesis and findings. An overall judgement of the risk of bias for a systematic review is also made. Each of the seven systematic reviews were assessed by two independent assessors using the ROBIS tool. A third and fourth assessor resolved any disagreements by consensus (see supplementary material for further information). As we demonstrate for each of the domains described below, our analyses resulted in all of the systematic reviews being judged as at a high risk of bias due to both methodological limitations and failure to adequately address these limitations in their conclusions and interpretations.

1.1. Domain 1: Study eligibility criteria

In this domain, all the systematic reviews were considered to be at high risk of bias because the eligibility criteria in most of the systematic reviews are unacceptably ambiguous, there are significant deviations from the eligibility criteria in the protocol, and the eligibility criteria excluded types of studies that are relevant to the systematic reviews' research questions. These problems are exacerbated by the fact that the protocol was updated on January 23rd 2023 to record that screening against the eligibility criteria had been completed (presumably for all of the systematic reviews) without taking this opportunity to record and explain deviations from the original criteria. Best practice guidelines for systematic reviews require that deviations like this are described and justified¹³.

Every systematic review, except one focusing on clinical guidelines, excluded all grey literature (in another deviation from the protocol) and all literature not in English. The exclusion of "grey" literature (including dissertations, white papers, and government reports) and non-English materials creates concerns that studies relevant to the systematic reviews may have been excluded. While it is true that much of the known scientific literature on this topic is published in English, the Cass Report's team had the opportunity and resources to search for data without language limitations given the support the review process received. If the goal was to conduct a thorough overview of all extant knowledge on the subject, these limitations—apart from being incongruent with best practices in the absence of justification—obstruct that goal.

The exclusion of qualitative research from the overall set of systematic reviews is questionable, especially because qualitative studies were part of the inclusion criteria in the protocol and several of the research questions for the reviews have been investigated using rigorous qualitative methods¹⁴⁻¹⁷. Given that the Cass Report itself presents qualitative data (anecdotes, quotations, community claims) as evidence and purports to value stakeholder input, the fact that qualitative studies were not reviewed impedes the comprehensiveness of the overall project.

1.2. Domain 2: Identification and selection of studies

For the second domain, all seven systematic reviews were judged to be at high risk of bias since it is likely that the search strategy failed to identify all relevant studies. A single search strategy was used for all systematic reviews without any modifications. This calls into question the appropriateness of the search strategy and its applicability to each specific systematic review. Furthermore, there is no evidence that grey literature was searched for, apart from the systematic review of clinical guidelines (despite the protocol indicating intent to do so).

For some reviews, the selection and identification of studies for inclusion was questionable. The systematic review for social transition excluded studies in which social transition was not treated "as an exposure", and therefore excluded Olson et al. (2018)¹⁸ and Rae et al. (2019)¹⁹. However, the authors included five other studies from the same ongoing project (the TransYouth Project) ignoring that the same design limitations would also apply to these studies, especially as no differences between the transgender group and the cisgender controls in this group would indicate something very different than no difference between a socially transitioned group and a group denied social transition.

1.3. Domain 3: Data collection and study appraisal

In the third domain, the systematic review of clinical guidelines was judged to have a low risk of bias, and each of the other systematic reviews was judged to be at a high risk of bias. This was due to concerns about whether the inclusion of studies for synthesis in each systematic review depended on how well these studies were reported, as no information was provided about how or whether missing data was sought from study authors, and problems with the application, or lack, of study appraisal.

There were several issues with how study appraisal was conducted in these reviews. In the systematic review of psychosocial support interventions for children and adolescents experiencing gender dysphoria or incongruence, the Mixed Methods Appraisal Tool²⁰ (MMAT) was used to critically appraise the quality of the primary studies. The reviewers categorised the study quality into three categories (low/medium/high), which is not recommended by the authors of the MMAT. In fact, the authors of the MMAT actively discourage the use of an overall score of quality and state that, if this advice is not followed, a 5-point scale should be used alongside a full description of the MMAT results.²¹

In the systematic review of clinical guidelines, which was not described in the protocol, an appropriate study appraisal tool—the AGREE-II—was used. However, its reliability in this context may be questionable since several other systematic reviews have applied this tool to some of the same studies and arrived at quite different conclusions, with some also criticising the usefulness of AGREE-II for GAC guidelines.^{22–24} Notably, the most restrictive guidelines were rated highest in this systematic review, without a clear justification related to how these guidelines were developed,^{6,25} while guidelines for more affirmative models of care were judged to be of higher quality in other reviews than by Taylor and colleagues.^{22–24}

An adapted version of the Newcastle-Ottawa Quality Assessment Scale²⁶ (NOS) was used in three of the systematics instead of the MMAT^{2,3,7}—a deviation from the

protocol that was not explained or clearly reported. The NOS has been highly criticised²⁷ and the use of an adapted scoring negates any previous attempts to validate the NOS. In fact, one of the systematic reviews⁷ cites this critical paper by Stang and colleagues (2010)²⁷ to support their use of the NOS—a practice its authors that have called out as a major quotation error.²⁸ The systematic review authors provide no rationale for the threshold scores used to categorise the quality of studies. It is generally accepted that the use of a single score (as used in these systematic reviews) is unacceptable in the assessment of risk of bias of individual studies within systematic reviews.^{29,30} Single scores do not capture the nuances of risk of bias, making them difficult to interpret. Additionally, the NOS is considered a quality appraisal scale, but within systematic reviews exploring the effectiveness of interventions (which this review claims to be doing) it is recommended that a domain-based risk of bias assessment tool should be used instead of a quality appraisal tool.¹¹ The ROBINS-I is an example of an available tool which is more suitable.³¹

1.4. Domain 4: Synthesis and findings

All of the systematic reviews were judged to be at high risk of bias for the synthesis and findings domain. For two of the systematic reviews, this was because of the inappropriate exclusion from the synthesis of studies that were deemed to be of "low quality" according to the adapted NOS.^{2,8} Using this approach, the authors excluded 48% and 36% of studies for puberty blockers and hormone therapy respectively. This practice is not recommended in systematic reviews unless explicitly pre-specified in a protocol with a clear and reasonable rationale, which was not evident in these cases.³² Instead, the narrative syntheses should have included all studies and integrated observations regarding study quality into the analyses, similar to how a sensitivity analysis would treat study quality in a meta-analysis.³³ In the systematic reviews on the characteristics of children and young people referred for GAC and on care pathways for this population, there was no assessment of study quality or risk of bias so their conclusions could not take these issues into account.

This pattern of deviations from the protocol's plan for quality assessment is striking. The protocol stated that the MMAT would be used to appraise the quality of the studies included in each systematic review. However, only one of the systematic reviews followed the protocol by using the MMAT, but did so inappropriately; the systematic review of clinical guidelines used an appropriate tool for quality assessment, but was not mentioned in the protocol; three of the systematic reviews used a different tool from what was planned in the protocol and altered it in problematic ways; and two of the systematic reviews did not assess study quality at all. It is notable that the combination of using the NOS instead of the MMAT, altering how it is scored, and then excluding evidence on the basis of this altered score only applied to the systematic reviews on what could be considered the three most controversial topics that the Cass Report addressed—puberty blockers, hormone therapy, and social transition. The fact that these decisions were deviations from the protocol and that justifications for them were not provided raises concerns about cherry-picking.

In the synthesis sections, the authors conclude that low quality or even moderate quality evidence indicates that there is insufficient evidence for recommending gender-affirming care practices. For example, in the review on hormone replacement therapy,

the authors argue that no conclusions can be drawn regarding any relationships between HRT and psychological health. However, this minimizes their own data. One study showed an improvement in gender dysphoria,³⁴ one showed an improvement in body satisfaction,³⁵ four studies showed an improvement in depression-related outcomes,^{34–37} three studies showed an improvement in anxiety-related outcomes^{34–36} and three studies showed a decrease in suicidality/self-harm related outcomes.^{36–38} Here, five distinct studies on 415 trans youth show an improvement in mental health, body satisfaction and/or gender dysphoria. Further, no study demonstrated “consistent” evidence for harm. Therefore, it appears that the body of evidence, despite its limitations, is at least suggestive of a beneficial direction. At minimum, the lack of evidence for harm should have been detailed. Had this been detailed more, the authors still would have likely concluded that more research is necessary for elucidating the impacts of gender-affirming care, but their conclusions would be less likely to be misrepresented to insinuate that GAC is harmful.

1.5. Additional Issues of Concern

Beyond what is captured by our application of ROBIS, there are several other ways in which these systematic reviews deviate from best practice. PRISMA reporting guidelines advise authors to provide a transparent, complete and accurate account of why a review was done, what they did, and what they found. However, as highlighted above, there are several deviations from PRISMA within the systematic reviews.³⁹ Of particular concern is the lack of explicit, pre-specified methods in a detailed protocol for each separate review resulting in a lack of transparency and reproducibility. The protocol falls short of the PRISMA-P guidelines⁴⁰ by failing to document changes to the protocol or how they would be recorded (item 4), specifying ambiguous eligibility criteria (item 8), failing to disclose how the search strategy was developed and the expertise of the searcher (item 10), omitting a description of whether or how information missing from included studies would be obtained (item 11), and neglecting to consider how meta-biases (e.g., publication bias, outcome reporting bias) may have affected the reviews (item 16). Ideally, an adequately detailed protocol should have been written for each individual systematic review and submitted for peer review.

Another deviation from best practice concerns the composition of the review team. Both the Cochrane Handbook¹¹ and the Institute of Medicine⁴¹ recommend including content area experts on the review team. Initially, the Cass team specifically excluded content experts, but they later added Dr Trilby Langton who is “a former Clinical Psychologist at the Tavistock Gender Identity Development Service”.⁹ Despite this, there is still a distinct lack of content expertise among the authors on many of the issues examined by the systematic reviews. Ideally, there would also be input to systematic reviews from those affected by the topic, and there is evidence that this practice is becoming common.⁴² Given the potential, and now actual, implications of the Cass Report for the gender-diverse communities in the UK and elsewhere, knowledge and experience from community leaders should have formed the basis for each of the systematic reviews.

Finally, it is considered mandatory by Cochrane to rerun searches that are more than 12 months old and to screen the results for eligibility. The search for these systematic reviews was conducted in May 2021 and updated in April 2022 and was therefore 24

months old on publication. This means that the most recent eligible research is not included in these reviews.

Not only were these systematic reviews out of date and conducted in a manner that is likely to have biased their conclusions, but their necessity is also questionable in some cases. Several previous studies had addressed similar research questions regarding puberty suppression and hormone therapy, for example, and by systematically reviewing the same evidence came to more positive conclusions regarding their usefulness for gender diverse children and young people.^{43–45} Next, we consider the new evidence that was generated for the Cass Report.

2. The Primary Research

The Cass Report commissioned primary research consisting of (i) a retrospective cohort study using electronic primary care records, (ii) an international survey of gender services for children and young people, and (iii) a qualitative study. Despite the methodological flaws in the cohort study and the qualitative study, which we demonstrate below, these studies were cited in the Cass Report, without due attention to its limitations, to support several of its claims (see for example 10.70 and 10.71, p. 146).

In the study of electronic health records, the authors aimed "to estimate for people aged 18 and under with GD: changes in incidence and prevalence over time". They did not account for changing acceptance, stigma, diagnostic criteria and clinical coding available to GPs (i.e. from DSM-IV to DSM-V and ICD-10 to ICD-11),^{46,47} or clinical guidelines, which may alter sampling over time and bias GD prevalence estimates (see "chronology bias" or "surveillance bias").^{48,49} The authors fail to demonstrate that the observed increase is either unexpected or of concern, yet the assumption of both underpins the Cass Review and its commissioning. Regarding the co-occurrence of ASD and GD, the authors conclude that this has increased, without appropriate statistical tests (e.g., time trend analysis)⁵⁰ or consideration of changes in the visibility and diagnosis of ASD, despite also warning of large confidence intervals. The authors also claim there was a two-phase growth in referrals for GAC, with an "acceleration" in 2015, without justifying this by statistically modelling the claim and comparing it to alternative models of the data. In the Cass Report, a discussion of statistical significance, bias, type I errors, and power is included (p. 51), as well as description of statistical significance (and lack thereof) in cited studies (p. 187). Given this evidence of statistical understanding, the failure to apply the same critical thought to the results of the primary research suggests a potential double standard.

The qualitative research summary links to a study protocol submitted before the research began. Critically, there is a lack of appropriate information about the overall methods, especially recruitment methods, and no discussion of selection biases (also lacking in Hall et al., 2024⁷). Additionally, there is no mention of reflexivity or positionality (where the researcher stands in relationship to those they are interviewing), which has consequences for the data interpretation and is an important indicator of rigour in qualitative research of healthcare.⁵¹ The lack of consideration for how rigorous qualitative studies are designed, combined with the aforementioned exclusion of qualitative research from the systematic reviews, and the use of single (and sometimes

misrepresented; see below) quotes from participants to support much broader conclusions demonstrates a misunderstanding of qualitative evidence in the Cass Report and the valuable insights about GAC that have been generated through qualitative research (see Horton, 2023⁵²).

As outlined above, the results from these studies carried out for the Cass Report were used to produce conclusions and recommendations about GAC without sufficient warning of their methodological limitations, in stark contrast to the exclusion of research with far fewer limitations from the systematic reviews. This seeming double standard calls into question the Cass Review's claims that it has been "[s]ystematically reviewing and evaluating the evidence" (p. 65, our emphasis).

3. How the Cass Report frames GAC

These flaws have not prevented proponents of the Cass Report from praising its application of EBM whilst criticising existing literature on GAC as "substandard".⁵³ This view is based upon an inappropriate use of a paternalistic lens which views GAC as quasi-psychiatric care, and upon an inappropriate methodological lens which centres randomised controlled trials and downplays the value of high-quality observational data, issues that we turn to next as they are also apparent in the Cass Report.

3.1. Use of a paternalistic lens

Recognising and supporting the authenticity and competence of transgender young people is an important aspect of the provision of high-quality care.⁵⁴ However, the Cass Report emphasises their distress, rather than their treatment wishes: the report describes them as "children with gender dysphoria and/or gender-related distress" (p.52) and then emphasises the resolution of this distress as the main goal of interventions. Framed in this way, GAC becomes one of several treatment options for a quasi-psychiatric condition, rather than the authentic preference of competent individuals (note that Gillick competence is still applicable, without special limits, to under-16s seeking GAC, after the Court of Appeal quashed the High Court judgement that set restrictions specifically for GAC).⁵⁵ The reviewers' approach allows them to consider alternatives which they allege are in equipoise with GAC due to a lack of evidence, but which run contrary to patient wishes.⁵⁶ Given that transgender people have a care need rather than a disease and seek actualisation of their identities as opposed to a cure, this paternalistic lens is inappropriate.⁵⁷ Moreover, such a lens is also generally inappropriate in psychiatric care, where patient autonomy should be supported wherever possible.⁵⁸

GAC should instead be considered through a similar lens as reproductive healthcare, akin to how healthcare providers and the public think about contraception, hormone replacement therapy, or fertility treatment.⁵⁷ Reproductive care requires not just the absence of illness, but "a state of physical, emotional, mental, and social well-being in relation to all aspects of sexuality and reproduction".⁵⁹ All individuals have the right to make decisions regarding their own reproductive care and must have access to services that support that right. Having a young person with GD undergo their natal puberty is not a neutral or desirable act just because it is a natural occurrence, in the same way that continuing an unwanted pregnancy or having intrusive menopausal symptoms

should not be considered the default option. By failing to use a reproductive healthcare lens, the review risks creating an environment where non-affirming alternatives can be undertaken contrary to competent patients' wishes, where unethical controlled studies can be performed (see below), and in which the role of observational and cohort studies is downplayed.⁵⁶

3.2. Use of an inappropriate methodological standard

The Cass Report sets out that randomised controlled trials (RCTs) are the gold standard to assess the efficacy of gender-affirming care (GAC), leading to the implication that the "research protocol" which will be the only method of accessing puberty blockers will be an RCT (e.g., p. 177). Indeed, the benefit of RCTs lies in their high internal validity, achieved through the randomisation process which reduces biases related to confounding factors.⁶⁰ However, their external validity is often criticized,⁶¹ both regarding study population representation and extrapolating experimental conditions to real-life settings.⁶² In transgender medicine, youths willing to subject themselves to a RCT are likely not representative of the broader population, and the homogeneous treatments in an RCT do not easily translate to individualized care in clinical practice.⁵⁶ Perhaps of most concern is the unethical and coercive nature of access to puberty blockers being contingent on consenting to participation in research.

Moreover, the causal agent in an RCT comes from the contrast between treatment and control groups to determine a treatment's effect.⁶² In GAC, belonging to the control group, rather than the absence of treatment, can affect the outcome, threatening internal validity. Blinding is impossible due to the obvious effects of puberty blockers or hormone treatments, likely causing control group participants to feel resentful demoralization.⁵⁶ This resentment can bias responses of the controls or lead them to self-destructive behaviours,⁵⁶ shifting the causal link from the treatment effect to the knowledge of group assignment. Differential attrition is also likely, as youths with supportive families, better socioeconomic status, or living in areas with a better availability of GAC may leave the study if in the control group or not participate at all.⁵⁶

Additionally, hormone treatments take time to show effects,⁶³ and various interventions may be needed at different times based on individual needs. This requires long-term follow-up in RCTs. However, the benefits of randomization diminish over time, leading to biases similar to observational studies.⁶⁰ This is especially true in transgender medicine, where affirmed youths' life trajectories differ from those without access to GAC, due to the experience of living in their authentic gender and differential exposure to discrimination. Participants might also access other types of GAC at different rates, adding more confounding factors. Thus, over the necessary duration to assess hormone treatment efficacy on wellbeing, the groups would likely diverge enough to lose the benefits of randomization, reducing the RCT's internal validity to that of a well-conducted observational study, which would not present the same ethical issues.

Finally, and perhaps more fundamentally, evaluating the efficacy of GAC based on psychosocial well-being is misguided. The primary goal of GAC is to prevent or induce the appearance of certain physical characteristics, and their physiological efficacy is undisputed. Mental health benefits are a logical consequence of living authentically.⁵⁷ Advocating for RCTs with mental health outcomes frames transness as a quasi-psychiatric condition, a distress whose suffering must be alleviated by the most

evidenced-based methods, which contradicts the depathologisation of transness and its recognition as an issue of bodily autonomy and human rights.^{64,65} Improvement in well-being does not come from a physiological action of hormones, which could be adequately isolated by an RCT, but from a combination of factors contributing to increased congruence. Proposing RCTs with a mental health outcome thus shows, at best, a profound misunderstanding of transness.

In their response to criticisms following the Review, the Cass team attempted to justify their criticism of cohort studies: "the same level of rigour should be expected when looking at the best treatment approaches for this population as for any other population so as not to perpetuate the disadvantaged position this group have been placed in when looking for information on treatment options".⁶⁶ The Cass team fail to mention that the majority of strong treatment recommendations in healthcare are based upon low or very low-quality evidence⁶⁷ or that, as elaborated above, cohort studies may be best suited for producing evidence that can best inform claims about GAC.

4. Unsupported Claims

Despite its emphasis on being evidence-based, the Cass Report includes many insufficiently evidenced claims that are used to inform conclusions and recommendations. We briefly describe select examples here (but for more, see Grijseels, 2024⁶⁸).

In an analysis of changes in patient profile, the authors stated that "the exponential increase in numbers within a 5-year timeframe is very much faster than would be expected for the normal evolution of acceptance of a minority group" (p. 118) but provided neither references nor data to support this claim, nor proper consideration of the complex relationship between referrals and factors associated with social acceptance.⁶⁹ Further, the authors made use of several flawed datasets and analyses, including double counting in referral data (Fig 11., p. 85), referral trajectories that are over 7 years old (Fig 15., p. 88), personal communications without associated methodology (Fig 16., p. 89), and flawed original research (see section 2). The authors also claimed that "the switch from birth-registered males to birth-registered females" is "unlike trans presentations in any prior historical period" (p. 26), without evidence for this change occurring, and without considering studies that contradict this claim.⁷⁰ On detransition, the authors claimed that there is a "suggestion that the numbers are increasing" (p. 33), again without demonstrating or citing evidence that this is happening while also failing to cite major studies on detransitioning (e.g., Turban et al., 2021⁷¹).

On social transition, the authors state that "others consider that it makes it more likely that a child's gender dysphoria, which might have resolved at puberty, has an altered trajectory potentially, culminating in life-long medical intervention" (p. 31). Despite their own review concluding that it is difficult to assess the impact of social transition due to the small volume and low quality of current research⁷ and that young people report reduced gender dysphoria and feeling more comfortable in themselves after socially transitioning (p.159), recommendations centre around considering partial rather than full transitioning for prepubertal children to prevent an altered "developmental trajectory". In addition to this contradiction, these recommendations frame early social transitioning and detransitioning through a lens of pathologisation that

leaves little room for the possibility that the formation of gender identity is non-linear. Such identity explorations during childhood may be experienced positively.⁷²

Finally, it appears the Cass report may have misrepresented participant quotes from its primary research to support its claims. For example, one participant is quoted as saying "there's not only one route or one set way to transition or be trans. They might want just hormones, or just surgery, people are different with different experiences, presentations, and bodies. It's fine for that to be the case, it's okay to have different plans for your medical transition." (p.147). While it seems that this participant is advocating for increased availability of care options for trans people, this quote directly informs 10.81, which suggests "it is important to inform people that medical transition is not the only option and that choosing not to go down that route does not invalidate their identity" (p.147). The report then reframes this response as evidence for the need to reduce the number of medical transitions, whereas this misinterpretation of what the person has said may reflect a confirmation bias from the review team that may have violated their participants' informed consent for research participation.

5. Political Context of the Cass Report

A further area of the Cass Report that requires scrutiny is its failure to comprehensively consider the wider sociopolitical context surrounding transgender healthcare, in the UK and worldwide. Systematic reviews should consider the social, cultural, and political context in which interventions are being implemented, especially when the intervention can affect the health of disadvantaged groups.⁷³ These contexts can influence both the outcome of interventions included within systematic reviews, and the implementation by policymakers of recommendations from reviews.^{73,74} However, in both the Cass Report itself and the systematic reviews upon which it is based, there are relatively few reflections on social and political context. Within the final report of the Review, the main contextual observation is that there are "polarised debates about a range of societal issues involving transgender people in the UK" (p. 67). The report, however, fails to adequately consider the evidence of significant increases in societal transphobia in recent years. For instance, the number of recorded hate crimes against transgender people reached a record high in 2023, while the British Social Attitudes survey shows a marked decline in attitudes towards transgender people: 36% of people now describe themselves as prejudiced against transgender people, twice as many as in 2019.^{75,76} Recent comparative research has found anti-transgender sentiments being shared widely by media actors and politicians within UK, as well as in several other European countries.⁷⁷

While striving for the best evidence-based interventions and practice across healthcare, it is crucial to consider how these sociopolitical trends might influence the implementation of health interventions aimed at transgender people and the evaluation of what constitutes evidence, expertise or best practice. For example, a recent study examined the presentation of expertise and evidence within state hearings about a ban on gender-affirming care for children and adolescents in Arkansas. The authors identified a clear trend of misrepresentation and disinformation by actors positioning themselves as clinical experts.⁷⁸ The specific themes of misrepresentation and disinformation in both Arkansas and other states throughout the U.S. include insinuations that being

trans is, in and of itself, evidence of mental illness and can be "cured" through talk therapy, outdated and misleading claims about the process of gender-affirming care in the U.S., misrepresenting the evidence regarding "desistance" in transgender youth, and rejection of scientific and medical authority.⁷⁹ By making this point, we make no assessment or claim of political bias on the part of the authors of the Cass Report and associated systematic reviews. Rather, we suggest that it is not possible for these authors to separate their research from this social and political context, as they attempt to do. The quality of their recommendations and the weight given to them should be critically considered while taking this context into account.

6. The future application of EBM in Gender Affirming Healthcare

The Cass Report's editorial argues for the importance of EBM to support clinicians in working with the everyday concerns and unknowns of practice. Central to EBM are the three pillars of best available evidence, clinical expertise, and the values and preferences of those accessing care. It is helpful to consider what the best available evidence could look like, in an approach tailored to the context of GAC, the view and preferences of gender diverse children and young people, and those who support them, and the clinical expertise of healthcare providers who deliver GAC.

The consideration of values and preferences have been historically absent in trans care—with its continued exclusion rooted in a legacy of pathologisation. Even among other cohorts that remain pathologised, efforts are increasingly made to value "experts by experience", including in the development of clinical guidelines.⁸⁰ Effective co-production needs to involve the community at every stage, not just superficially, as is common in NHS England initiatives.⁸¹ In a good example of this, Ziegler led two reviews of clinical practice guidelines for adults and for children in primary care for which the broader team included both members of the trans community and primary care GAC providers.^{23,24}

As outlined above, the Cass Report does not consider all of the best available evidence regarding GAC for children and young people and applies generic standards of evidence rather than considering what is the best possible evidence in this context given methodological, practical, and ethical constraints. It is our view that the best possible evidence regarding GAC is produced when the engagement and trust of participants is maximised through community involvement and clear communication,^{82,83} when these participants are diverse and followed longitudinally in rigorous observational designs,^{56,84} and to measure outcomes that are considered important by gender diverse children and young people, and those who support them, including the clinical expertise of healthcare providers who deliver GAC, using culturally appropriate and valid measures.^{85–87} These recommendations are consistent with the methodological standards for validity, generalisability, and patient-centredness set out by the Patient-Centred Outcomes Research Institute⁸⁸ and capture the three pillars of EBM.

7. Conclusion

We have demonstrated that the Cass Report's application of EBM to GAC for children and young people is deeply flawed. Our critical analysis reveals significant

methodological problems in the commissioned systematic reviews and primary research that undermine the validity of the Cass Report's recommendations. In light of this, and the Cass Report's poor understanding of transgender identities and experiences, it is vital the academic community question the integrity and validity of the Review's recommendations.

To truly uphold the principles of EBM, future research on GAC must generate high-quality observational data, involve transgender communities, and prioritise patient-centred outcomes. This approach ensures the validity, generalisability, and cultural appropriateness of findings. Only by addressing these critical gaps can we ensure that the healthcare system provides the necessary support and recognition for gender-diverse youth, aligning with ethical standards and promoting their well-being and autonomy.

Acknowledgements

Thanks to Edward Cunningham-Oakes for providing epidemiological expertise, feedback, and discussion on the primary research.

References

- [1] NHS England, Terms of reference for review of gender identity development service for children and adolescents, <https://www.england.nhs.uk/publication/terms-of-reference-for-review-of-gender-identity-development-service-for-children-and-adolescents/> (2020).
- [2] J. Taylor, A. Mitchell, R. Hall, T. Langton, L. Fraser, C. E. Hewitt, Masculinising and feminising hormone interventions for adolescents experiencing gender dysphoria or incongruence: A systematic review, *Archives of Disease in Childhood* (Apr. 2024). doi:10.1136/archdischild-2023-326670.
- [3] J. Taylor, A. Mitchell, R. Hall, C. Heathcote, T. Langton, L. Fraser, C. E. Hewitt, Interventions to suppress puberty in adolescents experiencing gender dysphoria or incongruence: A systematic review, *Archives of Disease in Childhood* (Apr. 2024). doi:10.1136/archdischild-2023-326669.
- [4] J. Taylor, R. Hall, T. Langton, L. Fraser, C. E. Hewitt, Characteristics of children and adolescents referred to specialist gender services: A systematic review, *Archives of Disease in Childhood* (Apr. 2024). doi:10.1136/archdischild-2023-326681.
- [5] J. Taylor, R. Hall, T. Langton, L. Fraser, C. E. Hewitt, Care pathways of children and adolescents referred to specialist gender services: A systematic review, *Archives of Disease in Childhood* (Apr. 2024). doi:10.1136/archdischild-2023-326760.
- [6] J. Taylor, R. Hall, C. Heathcote, C. E. Hewitt, T. Langton, L. Fraser, Clinical guidelines for children and adolescents experiencing gender dysphoria or incongruence: A systematic review of recommendations (part 2), *Archives of Disease in Childhood* (Apr. 2024). doi:10.1136/archdischild-2023-326500.
- [7] R. Hall, J. Taylor, C. E. Hewitt, C. Heathcote, S. W. Jarvis, T. Langton, L. Fraser, Impact of social transition in relation to gender for children and adolescents: A systematic review, *Archives of Disease in Childhood* (Apr. 2024). doi:10.1136/archdischild-2023-326112.
- [8] C. Heathcote, J. Taylor, R. Hall, S. W. Jarvis, T. Langton, C. E. Hewitt, L. Fraser, Psychosocial support interventions for children and adolescents experiencing gender dysphoria or incongruence: A systematic review, *Archives of Disease in Childhood* (Apr. 2024). doi:10.1136/archdischild-2023-326347.
- [9] L. Fraser, R. Hall, J. Taylor, C. Heathcote, S. W. Jarvis, K. Atkin, T. Langton, C. E. Hewitt, The epidemiology, management and outcomes of children with gender-related distress / gender dysphoria: A systematic review (2021). arXiv:CRD42021289659.
- [10] P. Whiting, J. Savović, J. P. Higgins, D. M. Caldwell, B. C. Reeves, B. Shea, P. Davies, J. Kleijnen, R. Churchill, ROBIS: A new tool to assess risk of bias in

- systematic reviews was developed, *Journal of Clinical Epidemiology* 69 (2016) 225–234. doi:10.1016/j.jclinepi.2015.06.005.
- [11] JPT. Higgins, J. Thomas, J. Chandler, M. Cumpston, T. Li, MJ. Page, VA. Welch, *Cochrane Handbook for Systematic Reviews of Interventions*, Cochrane, 2023.
- [12] R. Perry, A. Whitmarsh, V. Leach, P. Davies, A comparison of two assessment tools used in overviews of systematic reviews: ROBIS versus AMSTAR-2, *Systematic Reviews* 10 (1) (2021) 273. doi:10.1186/s13643-021-01819-x.
- [13] B. T. Johnson, E. A. Hennessy, Systematic reviews and meta-analyses in the health sciences: Best practice methods for research syntheses, *Social science & medicine* (1982) 233 (2019) 237–251. doi:10.1016/j.socscimed.2019.05.035.
- [14] K. Cooper, A. Russell, W. Mandy, C. Butler, The phenomenology of gender dysphoria in adults: A systematic review and meta-synthesis, *Clinical Psychology Review* 80 (2020) 101875. doi:10.1016/j.cpr.2020.101875.
- [15] R. Goulding, J. Goodwin, A. O'Donovan, M. M. Saab, Transgender and gender diverse youths' experiences of healthcare: A systematic review of qualitative studies, *Journal of Child Health Care* (2023) 13674935231222054doi:10.1177/13674935231222054.
- [16] R. S. Jessen, I. R. H. Haraldsen, E. Stänicke, Navigating in the dark: Meta-synthesis of subjective experiences of gender dysphoria amongst transgender and gender non-conforming youth, *Social Science & Medicine* 281 (2021) 114094. doi:10.1016/j.socscimed.2021.114094.
- [17] S. Kearns, T. Kroll, D. O'Shea, K. Neff, Experiences of transgender and non-binary youth accessing gender-affirming care: A systematic review and meta-ethnography, *PLOS ONE* 16 (9) (2021) e0257194. doi:10.1371/journal.pone.0257194.
- [18] K. R. Olson, S. Gülgöz, Early Findings From the TransYouth Project: Gender Development in Transgender Children, *Child Development Perspectives* 12 (2) (2018) 93–97. doi:10.1111/cdep.12268.
- [19] J. R. Rae, S. Gülgöz, L. Durwood, M. DeMeules, R. Lowe, G. Lindquist, K. R. Olson, Predicting Early-Childhood Gender Transitions, *Psychological Science* 30 (5) (2019) 669–681. doi:10.1177/0956797619830649.
- [20] P. Pluye, M.-P. Gagnon, F. Griffiths, J. Johnson-Lafleur, A scoring system for appraising mixed methods research, and concomitantly appraising qualitative, quantitative and mixed methods primary studies in *Mixed Studies Reviews*, *International Journal of Nursing Studies* 46 (4) (2009) 529–546. doi:10.1016/j.ijnurstu.2009.01.009.

- [21] Q. N. Hong, S. Fàbregues, G. Bartlett, F. Boardman, M. Cargo, P. Dagenais, M.-P. Gagnon, F. Griffiths, B. Nicolau, A. O’Cathain, M.-C. Rousseau, I. Vedel, P. Pluye, The Mixed Methods Appraisal Tool (MMAT) version 2018 for information professionals and researchers, *Education for Information* 34 (4) (2018) 285–291. doi:10.3233/EFI-180221.
- [22] M. I. Kruse, A. Clarizio, S. Karabelas-Pittman, B. L. Bigham, S. Upadhye, Systematic Review, Quality Assessment, and Synthesis of Guidelines for Emergency Department Care of Transgender and Gender-diverse People: Recommendations for Immediate Action to Improve Care, *Western Journal of Emergency Medicine* 25 (1) (Jan. 2024). doi:10.5811/WESTJEM.60632.
- [23] E. Ziegler, B. Carroll, E. Charnish, Review and Analysis of International Transgender Adult Primary Care Guidelines, *Transgender Health* 6 (3) (2021) 139–147. doi:10.1089/trgh.2020.0043.
- [24] E. Ziegler, E. Charnish, B. Carroll, L. Layman-Pleet, A Critical Review of Clinical Practice Guidelines for Care of Transgender and Gender Diverse Children and Youth for Use by Primary Care Practitioners, *Transgender Health* 7 (5) (2022) 397–406. doi:10.1089/trgh.2020.0185.
- [25] J. Taylor, R. Hall, C. Heathcote, C. E. Hewitt, T. Langton, L. Fraser, Clinical guidelines for children and adolescents experiencing gender dysphoria or incongruence: A systematic review of guideline quality (part 1), *Archives of Disease in Childhood* (Apr. 2024). doi:10.1136/archdischild-2023-326499.
- [26] G. Wells, B. Shea, D. O’Connell, J. Robertson, J. Peterson, V. Welch, M. Losos, P. Tugwell, The Newcastle-Ottawa Scale (NOS) for Assessing the Quality of Nonrandomized Studies in Meta-Analysis, in: *3rd Symposium on Systematic Reviews: Beyond the Basics*, Oxford, UK, 2000, p. 15.
- [27] A. Stang, Critical evaluation of the Newcastle-Ottawa scale for the assessment of the quality of nonrandomized studies in meta-analyses, *European Journal of Epidemiology* 25 (9) (2010) 603–605. doi:10.1007/s10654-010-9491-z.
- [28] A. Stang, S. Jonas, C. Poole, Case study in major quotation errors: A critical commentary on the Newcastle-Ottawa scale, *European Journal of Epidemiology* 33 (11) (2018) 1025–1031. doi:10.1007/s10654-018-0443-3.
- [29] P. Herbison, J. Hay-Smith, W. J. Gillespie, Adjustment of meta-analyses on the basis of quality scores should be abandoned, *Journal of Clinical Epidemiology* 59 (12) (2006) 1249–1256. doi:10.1016/j.jclinepi.2006.03.008.
- [30] M. Viswanathan, M. T. Ansari, N. D. Berkman, S. Chang, L. Hartling, M. McPheeters, P. L. Santaguida, T. Shamliyan, K. Singh, A. Tsertsvadze, J. R. Treadwell, Assessing the Risk of Bias of Individual Studies in Systematic Reviews of Health Care Interventions, in: *Methods Guide for Effectiveness and Comparative Effectiveness Reviews*, AHRQ Methods for Effective Health Care, Agency for Healthcare Research and Quality (US), Rockville (MD), 2008, p. [Internet].

- [31] J. A. Sterne, M. A. Hernán, B. C. Reeves, J. Savović, N. D. Berkman, M. Viswanathan, D. Henry, D. G. Altman, M. T. Ansari, I. Boutron, J. R. Carpenter, A.-W. Chan, R. Churchill, J. J. Deeks, A. Hróbjartsson, J. Kirkham, P. Jüni, Y. K. Loke, T. D. Pigott, C. R. Ramsay, D. Regidor, H. R. Rothstein, L. Sandhu, P. L. Santaguida, H. J. Schünemann, B. Shea, I. Shrier, P. Tugwell, L. Turner, J. C. Valentine, H. Waddington, E. Waters, G. A. Wells, P. F. Whiting, J. P. Higgins, ROBINS-I: A tool for assessing risk of bias in non-randomised studies of interventions, *BMJ* 355 (Oct. 2016). doi:10.1136/bmj.i4919.
- [32] L. A. Harvey, M. P. Dijkers, Should trials that are highly vulnerable to bias be excluded from systematic reviews?, *Spinal Cord* 57 (9) (2019) 715–716. doi:10.1038/s41393-019-0340-y.
- [33] K. Kolaski, L. R. Logan, J. P. A. Ioannidis, Guidance to best tools and practices for systematic reviews, *Systematic Reviews* 12 (1) (2023) 96. doi:10.1186/s13643-023-02255-9.
- [34] D. López de Lara, O. Pérez Rodríguez, I. Cuellar Flores, J. L. Pedreira Masa, L. Campos-Muñoz, M. Cuesta Hernández, J. T. Ramos Amador, Psychosocial assessment in transgender adolescents, *Anales de Pediatría (English Edition)* 93 (1) (2020) 41–48. doi:10.1016/j.anpede.2020.01.004.
- [35] C. Grannis, S. F. Leibowitz, S. Gahn, L. Nahata, M. Morningstar, W. I. Mattson, D. Chen, J. F. Strang, E. E. Nelson, Testosterone treatment, internalizing symptoms, and body image dissatisfaction in transgender boys, *Psychoneuroendocrinology* 132 (2021) 105358. doi:10.1016/j.psyneuen.2021.105358.
- [36] R. Kaltiala, E. Heino, M. Työläjäarvi, L. Suomalainen, Adolescent development and psychosocial functioning after starting cross-sex hormones for gender dysphoria, *Nordic Journal of Psychiatry* 74 (3) (2020) 213–219. doi:10.1080/08039488.2019.1691260.
- [37] A. E. Green, J. P. DeChants, M. N. Price, C. K. Davis, 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) (2022) 643–649. doi:10.1016/j.jadohealth.2021.10.036.
- [38] L. R. Allen, L. B. Watson, A. M. Egan, C. N. Moser, Well-being and suicidality among transgender youth after gender-affirming hormones, *Clinical Practice in Pediatric Psychology* 7 (3) (2019) 302–311. doi:10.1037/cpp0000288.
- [39] M. J. Page, J. E. McKenzie, P. M. Bossuyt, I. Boutron, T. C. Hoffmann, C. D. Mulrow, L. Shamseer, J. M. Tetzlaff, E. A. Akl, S. E. Brennan, R. Chou, J. Glanville, J. M. Grimshaw, A. Hróbjartsson, M. M. Lalu, T. Li, E. W. Loder, E. Mayo-Wilson, S. McDonald, L. A. McGuinness, L. A. Stewart, J. Thomas, A. C. Tricco, V. A. Welch, P. Whiting, D. Moher, The PRISMA 2020 statement: An updated guideline for reporting systematic reviews, *BMJ* 372 (2021) n71. doi:10.1136/bmj.n71.

- [40] L. Shamseer, D. Moher, M. Clarke, D. Gherzi, A. Liberati, M. Petticrew, P. Shekelle, L. A. Stewart, Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015: Elaboration and explanation, *BMJ* 349 (2015) g7647. doi:10.1136/bmj.g7647.
- [41] Institute of Medicine (US) Committee on Standards for Systematic Reviews of Comparative Effectiveness Research, J. Eden, L. Levit, A. Berg, S. Morton (Eds.), *Finding What Works in Health Care: Standards for Systematic Reviews*, National Academies Press (US), Washington (DC), 2011.
- [42] Q. Zhou, H. He, Q. Li, J. Zhao, L. Wang, Z. Luo, X. Wang, Y. Chen, Patient and public involvement in systematic reviews: Frequency, determinants, stages, barriers, and dissemination, *Journal of Clinical Epidemiology* 170 (2024) 111356. doi:10.1016/j.jclinepi.2024.111356.
- [43] D. Chew, J. Anderson, K. Williams, T. May, K. Pang, Hormonal treatment in young people with gender dysphoria: A systematic review, *Pediatrics* 141 (4) (2018). doi:10.1542/peds.2017-3742.
- [44] G. Ramos, A. Mengai, C. Daltro, P. Cutrim, E. Zlotnik, A. Beck, Systematic Review: Puberty suppression with GnRH analogues in adolescents with gender incongruity, *Journal of Endocrinological Investigation* 44 (6) (2021) 1151–1158. doi:10.1007/s40618-020-01449-5.
- [45] L. Rew, C. Young, M. Monge, R. Bogucka, Review: Puberty blockers for transgender and gender diverse youth—a critical review of the literature, *Child and Adolescent Mental Health* 26 (1) (2021) 3–14. doi:10.1111/camh.12437.
- [46] M.-A. Crocq, How gender dysphoria and incongruence became medical diagnoses – a historical review, *Dialogues in Clinical Neuroscience* 23 (1) (2021) 44–51. doi:10.1080/19585969.2022.2042166.
- [47] E. Coleman, A. E. Radix, W. P. Bouman, G. R. Brown, A. L. C. de Vries, M. B. Deutsch, R. Ettner, L. Fraser, M. Goodman, J. Green, A. B. Hancock, T. W. Johnson, D. H. Karasic, G. A. Knudson, S. F. Leibowitz, H. F. L. Meyer-Bahlburg, S. J. Monstrey, J. Motmans, L. Nahata, T. O. Nieder, S. L. Reisner, C. Richards, L. S. Schechter, V. Tangpricha, A. C. Tishelman, M. A. A. Van Trotsenburg, S. Winter, K. Ducheny, N. J. Adams, T. M. Adrián, L. R. Allen, D. Azul, H. Bagga, K. Başar, D. S. Bathory, J. J. Belinky, D. R. Berg, J. U. Berli, R. O. Bluebond-Langner, M.-B. Bouman, M. L. Bowers, P. J. Brassard, J. Byrne, L. Capitán, C. J. Cargill, J. M. Carswell, S. C. Chang, G. Chelvakumar, T. Corneil, K. B. Dalke, G. De Cuypere, E. de Vries, M. Den Heijer, A. H. Devor, C. Dhejne, A. D’Marco, E. K. Edmiston, L. Edwards-Leeper, R. Ehrbar, D. Ehrensaft, J. Eisfeld, E. Elaut, L. Erickson-Schroth, J. L. Feldman, A. D. Fisher, M. M. Garcia, L. Gijs, S. E. Green, B. P. Hall, T. L. D. Hardy, M. S. Irwig, L. A. Jacobs, A. C. Janssen, K. Johnson, D. T. Klink, B. P. C. Kreukels, L. E. Kuper, E. J. Kvach, M. A. Malouf, R. Massey, T. Mazur, C. McLachlan, S. D. Morrison, S. W. Mosser,

- P. M. Neira, U. Nygren, J. M. Oates, J. Obedin-Maliver, G. Pagkalos, J. Patton, N. Phanuphak, K. Rachlin, T. Reed, G. N. Rider, J. Ristori, S. Robbins-Cherry, S. A. Roberts, K. A. Rodriguez-Wallberg, S. M. Rosenthal, K. Sabir, J. D. Safer, A. I. Scheim, L. J. Seal, T. J. Schoole, K. Spencer, C. St. Amand, T. D. Steensma, J. F. Strang, G. B. Taylor, K. Tilleman, G. G. T'Sjoen, L. N. Vala, N. M. Van Mello, J. F. Veale, J. A. Vencill, B. Vincent, L. M. Wesp, M. A. West, J. Arcelus, Standards of Care for the Health of Transgender and Gender Diverse People, Version 8, *International Journal of Transgender Health* 23 (sup1) (2022) S1–S259. doi:10.1080/26895269.2022.2100644.
- [48] C. J. Pannucci, E. G. Wilkins, Identifying and Avoiding Bias in Research:, *Plastic and Reconstructive Surgery* 126 (2) (2010) 619–625. doi:10.1097/PRS.0b013e3181de24bc.
- [49] S. Tancredi, S. Cullati, A. Chiolero, Screening and Surveillance Bias in Cancer, *Epidemiologia* 4 (2) (2023) 117–120. doi:10.3390/epidemiologia4020012.
- [50] J. W. Ely, J. D. Dawson, J. H. Lemke, J. Rosenberg, An Introduction to Time-Trend Analysis, *Infection Control and Hospital Epidemiology* 18 (4) (1997) 267–274. arXiv:30141214, doi:10.2307/30141214.
- [51] K. Jamie, A. P. Rathbone, Using theory and reflexivity to preserve methodological rigour of data collection in qualitative research, *Research Methods in Medicine & Health Sciences* 3 (1) (2022) 11–21. doi:10.1177/26320843211061302.
- [52] C. Horton, The importance of child voice in trans health research: A critical review of research on social transition and well-being in trans children, *International Journal of Transgender Health* 0 (0) (2023) 1–18. doi:10.1080/26895269.2023.2295381.
- [53] K. Abbasi, The Cass review: An opportunity to unite behind evidence informed care in gender medicine, *BMJ* 385 (2024) q837. doi:10.1136/bmj.q837.
- [54] L. Crosse, Respecting the free will, authenticity and autonomy of transgender youth, *Nursing Ethics* (2023) 09697330231180743doi:10.1177/09697330231180743.
- [55] *Bell v. Tavistock* (Sep. 2021).
- [56] F. Ashley, D. M. Tordoff, J. Olson-Kennedy, A. J. Restar, Randomized-controlled trials are methodologically inappropriate in adolescent transgender healthcare, *International Journal of Transgender Health* (2023) 1–12doi:10.1080/26895269.2023.2218357.
- [57] F. Ashley, Adolescent Medical Transition is Ethical: An Analogy with Reproductive Health, *Kennedy Institute of Ethics Journal* 32 (2) (2022) 127–171. doi:10.1353/ken.2022.0010.

- [58] J. Bergamin, J. Luigjes, J. Kiverstein, C. L. Bockting, D. Denys, Defining Autonomy in Psychiatry, *Frontiers in Psychiatry* 13 (2022) 801415. doi:10.3389/fpsy.2022.801415.
- [59] WHO, Sexual health, [https://www.who.int/teams/sexual-and-reproductive-health-and-research-\(srh\)/areas-of-work/sexual-health](https://www.who.int/teams/sexual-and-reproductive-health-and-research-(srh)/areas-of-work/sexual-health) (2006).
- [60] R. I. Horwitz, J. E. Abell, J. B. Christian, A. E. Wivel, Right Answers, Wrong Questions in Clinical Research, *Science Translational Medicine* 6 (221) (Jan. 2014). doi:10.1126/scitranslmed.3007649.
- [61] T. R. Frieden, Evidence for Health Decision Making — Beyond Randomized, Controlled Trials, *New England Journal of Medicine* 377 (5) (2017) 465–475. doi:10.1056/NEJMr1614394.
- [62] M. Wadhwa, T. D. Cook, The Set of Assumptions Randomized Control Trials Make and Their Implications for the Role of Such Experiments in Evidence-Based Child and Adolescent Development Research, *New Directions for Child and Adolescent Development* 2019 (167) (2019) 17–37. doi:10.1002/cad.20313.
- [63] W. C. Hembree, P. T. Cohen-Kettenis, L. Gooren, S. E. Hannema, W. J. Meyer, M. H. Murad, S. M. Rosenthal, J. D. Safer, V. Tangpricha, G. G. T’Sjoen, Endocrine Treatment of Gender-Dysphoric/Gender-Incongruent Persons: An Endocrine Society* Clinical Practice Guideline, *The Journal of Clinical Endocrinology & Metabolism* 102 (11) (2017) 3869–3903. doi:10.1210/jc.2017-01658.
- [64] Yogyakarta Principles, The Yogyakarta Principles, <https://yogyakartaprinciples.org/principles-en/> (2007).
- [65] European Union Agency for Fundamental Rights., *Fundamental Rights Report 2017.*, Publications Office, LU, 2017.
- [66] Hilary Cass, Final report – FAQs, <https://cass.independent-review.uk/home/publications/final-report/final-report-faqs/> (2024).
- [67] P. E. Alexander, L. Bero, V. M. Montori, J. P. Brito, R. Stoltzfus, B. Djulbegovic, I. Neumann, S. Rave, G. Guyatt, World Health Organization recommendations are often strong based on low confidence in effect estimates, *Journal of Clinical Epidemiology* 67 (6) (2014) 629–634. doi:10.1016/j.jclinepi.2013.09.020.
- [68] D. Grijseels, Biological and psychosocial evidence in the Cass Review: A critical commentary (Apr. 2024). doi:10.31235/osf.io/wjafd.
- [69] F. Ashley, Shifts in Assigned Sex Ratios at Gender Identity Clinics Likely Reflect Changes in Referral Patterns, *The Journal of Sexual Medicine* 16 (6) (2019) 948–949. doi:10.1016/j.jsxm.2019.03.407.

- [70] J. S. Morandini, A. Kelly, N. M. de Graaf, P. Carmichael, I. Dar-Nimrod, Shifts in demographics and mental health co-morbidities among gender dysphoric youth referred to a specialist gender dysphoria service, *Clinical Child Psychology and Psychiatry* 27 (2) (2022) 480–491. doi:10.1177/13591045211046813.
- [71] J. L. Turban, S. S. Loo, A. N. Almazan, A. S. Keuroghlian, Factors Leading to “Detransition” Among Transgender and Gender Diverse People in the United States: A Mixed-Methods Analysis, *LGBT Health* 8 (4) (2021) 273–280. doi:10.1089/lgbt.2020.0437.
- [72] L. Durwood, K. A. Kuvalanka, S. Kahn-Samuels, A. E. Jordan, J. D. Rubin, P. Schnelzer, A. H. Devor, K. R. Olson, Retransitioning: The experiences of youth who socially transition genders more than once, *International Journal of Transgender Health* 23 (4) (2022) 409–427. doi:10.1080/26895269.2022.2085224.
- [73] P. Tugwell, M. Petticrew, E. Kristjansson, V. Welch, E. Ueffing, E. Waters, J. Bonney, A. Morgan, E. Doohan, M. P. Kelly, Assessing equity in systematic reviews: Realising the recommendations of the Commission on Social Determinants of Health, *BMJ* 341 (2010) c4739. doi:10.1136/bmj.c4739.
- [74] P. Cairney, K. Oliver, Evidence-based policymaking is not like evidence-based medicine, so how far should you go to bridge the divide between evidence and policy?, *Health Research Policy and Systems* 15 (1) (2017) 35–s12961–017–0192–x. doi:10.1186/s12961-017-0192-x.
- [75] E. Clery, BSA 40: A liberalisation in attitudes?, Tech. rep., National Centre for Social Research, London (2023).
- [76] J. Flatley, Hate crime, England and Wales, 2022 to 2023 second edition, Tech. rep., UK Home Office, United Kingdoms (2023).
- [77] RESIST, The RESIST Project: National and transnational reports on the formation of anti-gender politics., Tech. rep., University College Dublin (2024).
- [78] J. Wuest, B. S. Last, Agents of scientific uncertainty: Conflicts over evidence and expertise in gender-affirming care bans for minors, *Social Science & Medicine* 344 (2024) 116533. doi:10.1016/j.socscimed.2023.116533.
- [79] M. McNamara, Q. McLamore, N. Meade, M. Olgun, H. Robinson, A. Alstott, A thematic analysis of disinformation in gender-affirming healthcare bans in the United States, *Social Science & Medicine* 351 (2024) 116943. doi:10.1016/j.socscimed.2024.116943.
- [80] M. Garbi, National Institute for Health and Care Excellence clinical guidelines development principles and processes, *Heart (British Cardiac Society)* 107 (12) (2021) 949–953. doi:10.1136/heartjnl-2020-318661.

- [81] C. Marshall, A. Zambeaux, E. Ainley, D. McNally, J. King, L. Wolfenden, H. Lee, NHS England Always Events® program: Developing a national model for co-production, *Patient Experience Journal* 6 (1) (2019) 154–165. doi:10.35680/2372-0247.1340.
- [82] N. Adams, R. Pearce, J. Veale, A. Radix, D. Castro, A. Sarkar, K. C. Thom, Guidance and Ethical Considerations for Undertaking Transgender Health Research and Institutional Review Boards Adjudicating this Research, *Transgender Health* 2 (1) (2017) 165–175. doi:10.1089/trgh.2017.0012.
- [83] A. Asquith, L. Sava, A. B. Harris, A. E. Radix, D. J. Pardee, S. L. Reisner, Patient-centered practices for engaging transgender and gender diverse patients in clinical research studies, *BMC Medical Research Methodology* 21 (1) (2021) 202. doi:10.1186/s12874-021-01328-4.
- [84] M. B. Deutsch, A. Radix, S. Reisner, What’s in a Guideline? Developing Collaborative and Sound Research Designs that Substantiate Best Practice Recommendations for Transgender Health Care, *AMA Journal of Ethics* 18 (11) (2016) 1098–1106. doi:10.1001/journalofethics.2016.18.11.stas1-1611.
- [85] S. J. Bowman, L. J. Casey, J. McAloon, B. M. Wootton, Assessing gender dysphoria: A systematic review of patient-reported outcome measures, *Psychology of Sexual Orientation and Gender Diversity* 9 (4) (2022) 398–409. doi:10.1037/sgd0000486.
- [86] R. Kamran, L. Jackman, A. Laws, M. Stepney, C. Harrison, A. Jain, J. Rodrigues, Developing feasible and acceptable strategies for integrating the use of patient-reported outcome measures (PROMs) in gender-affirming care: An implementation study, *PLOS ONE* 19 (4) (2024) e0301922. doi:10.1371/journal.pone.0301922.
- [87] E. A. Tebbe, S. L. Budge, Research With Trans Communities: Applying a Process-Oriented Approach to Methodological Considerations and Research Recommendations, *The Counseling Psychologist* 44 (7) (2016) 996–1024. doi:10.1177/0011000015609045.
- [88] PCORI, Methodological Standards and Patient-Centeredness in Comparative Effectiveness Research: The PCORI Perspective, *JAMA* 307 (15) (2012) 1636–1640. doi:10.1001/jama.2012.466.

Supplementary Table 1. Summary of ROBIS Evaluations.

Domain 1: study eligibility criteria							
Systematic Review	Did the review adhere to pre-defined objectives and eligibility criteria?	Were the eligibility criteria appropriate for the review question?	Were eligibility criteria unambiguous?	Were any restrictions in eligibility criteria based on study characteristics appropriate?	Were any restrictions in eligibility criteria based on sources of information appropriate?	Concerns regarding specification of study eligibility criteria	
1	N	PY	PN	N	N	High	
2	N	N	N	N	N	High	
3	N	Y	N	Y	PY	High	
4	PN	PY	PY	N	N	High	
5	PY	Y	N	PY	N	High	
6	N	N	PN	Y	PN	High	
7	N	PN	N	PN	PN	High	
Domain 2: identification and selection of studies							
	Did the search include an appropriate range of databases /electronic sources for published and unpublished reports?	Were methods additional to database searching used to identify relevant reports?	Were the terms and structure of the search strategy likely to retrieve as many eligible studies as possible?	Were restrictions based on date, publication format, or language appropriate?	Were efforts made to minimise error in selection of studies?	Concerns regarding methods used to identify and/or select studies	
1	N	Y	Y	N	Y	High	
2	N	PY	PN	N	PY	High	
3	Y	Y	Y	Y	Y	High	
4	N	PY	PN	N	PY	High	
5	N	N	Y	N	Y	High	
6	PN	Y	PN	N	Y	High	
7	PN	PY	PN	N	PN	High	
Domain 3: data collection and study appraisal							
	Were efforts made to minimise error in data collection?	Were sufficient study characteristics available for both review authors and readers to be able to interpret the results?	Were all relevant study results collected for use in the synthesis?	Was methodological quality formally assessed using appropriate criteria?	Were efforts made to minimise error in risk of quality assessment?	Concerns regarding methods used to collect data and appraise studies	
1	NI	Y	Y	N	PY	High	
2	PY	Y	PN	PN	Y	High	
3	PY	Y	Y	PY	Y	Low	
4	PY	PY	N	N	Y	High	
5	PY	Y	PN	N	N	High	
6	PY	Y	PN	N	Y	High	
7	PY	Y	PY	N	N	High	
Domain 4: synthesis and findings							
	Did the synthesis include all studies that it should?	Were all pre-defined analyses reported or departures explained?	Was the synthesis appropriate given the nature and similarity in the research questions, study designs and outcomes across included studies?	Was between-study variation (heterogeneity) minimal or addressed in the synthesis?	Were the findings robust, e.g. as demonstrated through funnel plot or sensitivity analyses?	Were biases in primary studies minimal or addressed in the synthesis?	Concerns regarding the synthesis and findings
1	Y	N	N	Y	N	N	High
2	Y	PN	N	Y	N	N	High
3	Y	N	Y	Y	PN	N	High
4	N	N	N	Y	N	N	High
5	NI	N	N	N	N	N	High
6	N	N	PY	Y	N	PN	High
7	PN	N	PY	PY	N	N	High

Overall

	Did the interpretation of findings address all of the concerns identified in Domains 1 to 4?	Was the relevance of identified studies to the review's research question appropriately considered?	Did the reviewers avoid emphasising results on the basis of their statistical significance?	Risk of bias in the review
1	N	N	Y	High
2	N	PN	NI	High
3	N	Y	Y	High
4	N	PN	NI	High
5	N	N	Y	High
6	N	PN	PY	High
7	N	PN	Y	High

Note: For ease of reading, we have assigned a number to each systematic review that was evaluated using the ROBIS as follows:

- 1: Impact of social transition in relation to gender for children and adolescents: a systematic review,
- 2: Psychosocial support interventions for children and adolescents experiencing gender dysphoria or incongruence: a systematic review,
- 3: Clinical guidelines for children and adolescents experiencing gender dysphoria or incongruence: a systematic review of guideline quality,
- 4: Interventions to suppress puberty in adolescents experiencing gender dysphoria or incongruence: a systematic review,
- 5: Characteristics of children and adolescents referred to specialist gender services: a systematic review,
- 6: Masculinising and feminising hormone interventions for adolescents experiencing gender dysphoria or incongruence: a systematic review,
- 7: Care pathways of children and adolescents referred to specialist gender services: a systematic review