



Quality of the World Professional Association for Transgender Health Guideline Standards of Care 8: An Appraisal Using the AGREE II Instrument

Yuan Zhang¹ · Dominika Januś² · Riittakerttu Kaltiala^{3,4,5} · Milla Karvonen^{3,4} · Jeffrey J. Ptak⁵ · Juan Jose Yepes-Nuñez^{6,7}

Received: 15 October 2024 / Revised: 9 December 2025 / Accepted: 10 December 2025

© The Author(s) 2026

Abstract

In 2022, the World Professional Association for Transgender Health (WPATH) released their guidelines, Standards of Care Version 8 (SOC8), which have been regarded as establishing standards for the management of transgender patients. To conduct a quality assessment of the WPATH's SOC8 using the Appraisal of Guidelines for Research and Evaluation (AGREE) II instrument, we recruited a diverse group of international health professionals as assessors, including six clinicians and two guideline methodologists, who were trained in the AGREE II tool and underwent a calibration exercise. They independently assessed six SOC8 chapters: adolescents, children, hormone therapy, mental health, primary care, and surgery. We summarized the ratings across assessors for all items to present the strengths and limitations of the guidelines. The evaluations identified that the SOC8 chapters had limitations in the AGREE II domains of applicability (ranging from 28% to 40% across chapters), editorial independence (ranging from 43% to 44% across chapters), and rigor of development (ranging from 39% to 47% across chapters). On a scale of 1 to 7 (7 = highest quality), the median overall quality score was 3.5 to 4 for all chapters. Two of eight assessors recommended using the guideline chapters, while three recommended use with modification, and three recommended not to use the reviewed guideline chapters due to methodological concerns. Evidence-based guidelines addressing the needs of transidentified people are urgently needed, but it is imperative that healthcare providers, researchers, and policymakers recognize and address the limitations present in these six SOC8 chapters.

Keywords Transgender · Gender dysphoria · Children · Adolescents · AGREE II

Dominika Januś, Riittakerttu Kaltiala, Milla Karvonen, Jeffrey J. Ptak, and Juan Jose Yepes-Nuñez have contributed equally to this work and we ordered them alphabetically by last names.

✉ Yuan Zhang
evidencebridge@gmail.com

¹ Evidence Bridge, 2482 Yonge Street, Unit 1343, Toronto, ON M4P 2H5, Canada

² Jagiellonian University Medical College, Krakow, Poland

³ Faculty of Medicine and Health Technology, Tampere University, Tampere, Finland

⁴ Department of Adolescent Psychiatry, Tampere University Hospital, Tampere, Finland

⁵ Independent Physician, Private Practice, Scottsdale, AZ, USA

⁶ School of Medicine, Universidad de los Andes, Bogota, D.C., Colombia

⁷ Fundacion Santa Fe De Bogota, Bogota, D.C., Colombia

Introduction

Evidence-based medicine ensures the integration of the current best available evidence in healthcare decision making (Haynes et al., 2002; Sackett et al., 1996). It is an inherent and inseparable part of medical professionalism and ethics, upholding principles such as respect for autonomy, beneficence, nonmaleficence, and justice (Varkey, 2021). Following an evidence-based medicine approach allows healthcare providers to deliver care based on the best available evidence to benefit patients, and avoid or minimize harms (Stone, 2018). In clinical practice, guidelines offer evidence-based recommendations to healthcare providers (Woolf et al., 1999). Trustworthy guidelines, developed through rigorous and systematic methodology, have the potential to improve patient health. Conversely, poorly developed guidelines risk compromising the care of patients and even causing harm (Bahtsevani et al., 2004).

Over the past two decades, there has been significant advancement in the methodology of guideline development (Alonso-Coello et al., 2016a, 2016b; Guyatt et al., 2008; Rosenfeld et al., 2013; Shiffman et al., 2008; Woolf et al., 1999). It is widely accepted that evidence-based guidelines must be underpinned by a systematic review and critical appraisal of evidence, weighing the benefits and harms of care options. The “Clinical Practice Guidelines We Can Trust” document by the Institute of Medicine outlines standards for developing reliable guidelines (Institute of Medicine Committee on Standards for Developing Trustworthy Clinical Practice Guidelines, 2011). These standards include transparency, management of conflicts of interest, the intersection of systematic review with guideline development, establishing evidence foundations for and rating strength of guideline recommendations, articulation of recommendations, external review, and updating (Institute of Medicine Committee on Standards for Developing Trustworthy Clinical Practice Guidelines, 2011). Meanwhile, to support the assessment of guidelines, researchers have developed tools such as the Appraisal of Guidelines for Research and Evaluation (AGREE) II Instrument (Brouwers et al., 2010a, 2010b). The AGREE II Instrument is recognized as the benchmark tool for such assessments.

The World Professional Association for Transgender Health (WPATH), a prominent organization in transgender healthcare, offers the Standards of Care (SOC) as guidelines for healthcare providers serving the gender diverse population (World Professional Association for Transgender Health, 2024). These guidelines are widely regarded as establishing standards for the clinical management of transgender patients (Human Rights Campaign Foundation, 2024; Leibowitz, 2023; Restar et al., 2024). In 2022, WPATH released the eighth edition, SOC8, which aims to “provide clinical guidance for health professionals to assist transsexual, transgender, and gender nonconforming people with safe and effective pathways,” citing a “slowly growing body of evidence supporting the effectiveness of early medical intervention” (Coleman et al., 2022). SOC8 includes 18 chapters on a variety of topics and is designed for a diverse audience, including healthcare providers, individuals and their families, and social institutions. The diagnosis and management of gender incongruence and/or gender dysphoria, conditions that may be experienced by transidentified children and adolescents, remains a subject of intense debate and division. Clinicians using established guideline assessment tools, such as the AGREE II instrument, can gauge the methodological rigor and reliability of the recommendations from a healthcare provider’s perspective. This evaluation is crucial not only because it includes contributions from independent healthcare providers but also because it may further advance

evidence-based decision making and the development of rigorous guidelines in this sensitive and evolving field.

In this study, we engaged a diverse group of international healthcare providers and health research methodologists to conduct a quality appraisal of the WPATH’s SOC8 (Coleman et al., 2022). We assessed the chapters titled adolescents, children, hormone therapy, mental health, primary care, and surgery. These chapters were selected because they center on the diagnosis and management of transidentified children and adolescents.

Method

This study aimed to evaluate the guidelines from the perspectives of health professionals impacted by SOC8 recommendations. To minimize subjectivity in the assessment, this study included conflict of interest management, attitude screening before recruitment, as well as methodology training and calibration exercises for assessors after recruitment.

Eligibility and Recruitment of Assessors

We recruited a group of health professionals who collectively represented diverse rather than fixed views toward care models for transidentified children and adolescents, as well as minimal financial, intellectual, or professional conflicts of interest as assessors.

To ensure that diverse views were represented among those involved in the care of trans-identified children and adolescents, we invited healthcare providers including a range of specialists from family physicians to endocrinologists and pediatricians, with no geographic limitations. We sent invitations to potential assessors via contacts of the study lead methodologist and through university mailing lists (Supplementary material 1). Professionals who were interested in participating were briefed on the study objectives and expected commitment. For their 40-h expected contribution, assessors received compensation of \$1,500 USD from the research project.

We screened potential assessors’ disclosures for financial, intellectual, and professional conflicts. We adapted and refined a comprehensive form originally created by the British Medical Journal, which we then pilot-tested (Supplementary material 2) (Siemieniuk et al., 2016). Financial conflicts were defined as any monetary ties to entities with vested interests in the guideline’s subject matter, such as grants received or stock ownership (Guyatt et al., 2010; Hansen et al., 2019). Intellectual conflicts were any scholarly activities that might predispose an individual to a particular viewpoint, potentially compromising their impartiality regarding specific recommendations, exemplified by previous publications on SOC8-related topics (Guyatt et al., 2010).

Professional conflicts were defined as the assessors' professional engagements that intersect with the guideline's scope, such as clinical practices reliant on the interventions discussed in guidelines (Guyatt et al., 2010). Thus, having their practice consisting primarily of trans-identified children and adolescents is considered a professional conflict of interest. We excluded healthcare providers and methodologists with financial conflicts. We chose to limit rather than completely exclude people with potential professional conflicts of interest and a priori decided to include healthcare providers whose medical practice comprised no more than 25% in the management of trans-identified children and adolescents. Further, we limited people with intellectual and professional conflicts of interest to no more than 20% of assessors (see conflicts of interest criteria in Supplementary material 3). The investigator (YZ) reviewed the conflicts of interest screening form of potential assessors.

To minimize rigid beliefs and uncompromising viewpoints in the evaluation of the guidelines, we sought assessors who demonstrated a willingness to partake in constructive and respectful discourse. To this end, we introduced an attitude screening form, which encapsulated the study team's principles regarding the treatment of transgender children and adolescents. This form (Supplementary material 4) summarized what this study team considered reasonable principles about the evidence-based care of this population and asked assessors to agree with the following: (1) the management of transgender children and adolescents should be guided by the best available research evidence, incorporating patient values and preferences, and clinician expertise; (2) if there is compelling research evidence that gender-affirming medical interventions cause significant harms, medical professionals may need to discontinue providing these interventions for patients for whom these interventions may be more harmful than beneficial; and, on the other hand, if there is compelling research evidence that gender-affirming medical interventions are beneficial, medical professionals could prescribe these interventions more confidently; and (3) that when the evidence supporting a medical intervention is of low certainty, patients and clinicians should engage in shared decision making but that shared decision making is challenging with minors.

From March to October 2023, we invited 13 healthcare providers who work with trans-identified children and adolescents. Three declined due to time constraints, and one due to the politically sensitive nature of the topic. Of the nine willing participants, two were excluded for conflicts of interest, and one withdrew after methodology training, citing concerns about funding from the Society for Evidence-Based Gender Medicine and reputational risk. In total, six clinicians (pediatric endocrinologist, plastic surgeon, pediatrician, general physician, psychologist, and psychiatrist) completed training and assessed SOC8. We

further recruited two guideline methodologists who have more than ten years of guideline development experience as methodologist assessors.

Application of AGREE II to SOC8

The AGREE II tool is an internationally accepted tool and framework for evaluating the quality of guidelines, with 23 items across six domains: scope and purpose, stakeholder involvement, rigor of development, clarity of presentation, applicability, and editorial independence (Brouwers et al., 2010a, 2010b). Each domain of the AGREE II tool consists of two to nine items with seven response options (1 = Strongly disagree; 7 = Strongly agree with a statement about the quality of the guideline being assessed). To appraise guidelines, we trained assessors to use the AGREE II tool and applied this tool for clinical practice guidelines unrelated to gender-affirming care, as a calibration exercise (Supplementary material 5).

After training and calibration, assessors independently applied the AGREE II tool to the six chapters of the SOC8 that present guideline recommendations related to the diagnosis and management of gender dysphoria in children and adolescents. The Excel evaluation form was used by all assessors to ensure consistency in data collection. Each sheet contained 23 standardized AGREE II items and two overall assessment questions (one on the overall quality of the guideline and another on whether the guideline would be recommended for use in practice).

For each AGREE II item, the overall score, and whether to recommend the guideline for use, we allowed assessors to leave free-text comments addressing issues that may not have been captured by the AGREE II tool. After assessors submitted their assessment, one investigator (YZ) checked the completeness of the assessment and asked assessors to complete any missing values.

Synthesis and Analysis

We reported mean and median overall quality scores (ranging from 1 to 7) and calculated domain-specific scores (ranging from 0 to 100%) for the six AGREE II domains for each guideline chapter. According to the AGREE II user manual (Brouwers et al., 2010a, 2010b), domain scores are calculated using the formula

$$\frac{\text{Obtained score} - \text{minimum possible score}}{\text{Maximum possible score} - \text{minimum possible score}}$$

In this formula, the 'Obtained score' is the aggregate of all raters' scores for each item within a domain. The 'Minimum

score' and 'Maximum score' represent hypothetical scenarios where raters give all items in a domain the lowest (1) or highest (7) possible scores, respectively. Consequently, the domain score is expressed as a percentage of the maximum achievable score for that domain, ranging from 0 to 100%, with a higher score indicating better quality. We further summarized the mean and median ratings across assessors for all items. All analyses were performed using R (version 4.4.1) (RStudio Team, 2024).

We tabulated comments provided by the assessors in the overall assessment as the rationale for recommending or not recommending use of the guideline without any qualitative analysis or interpretation, preserving the original perspectives and insights of the assessors.

Results

Main Findings

Table 1 summarizes the domain-specific scores and overall quality score for each guideline chapter. The SOC8 chapters scored lowest for the applicability domain (scores between 28 and 40%) followed by the domain of rigor of development (39–47%) and editorial independence (43–44%). Meanwhile, the scores for scope and purpose (58–69%), stakeholder involvement (49–60%), and clarity of presentation (52–64%) were comparatively higher. On a scale from 1 to 7 (7 = highest quality), the median overall quality scores ranged from 3.5 to 4.0 across six chapters. The mean overall scores were similar: adolescents (3.4), children (3.5), hormone therapy (3.8), mental health (3.4), primary care (3.5), and surgery (3.5).

Table 2 provides details about the scores each assessor assigned to the AGREE II tool items for all six chapters. Supplementary material 6 summarizes the strengths and limitations revealed across AGREE II items. Notably, the median scores ranged between 1.5 and 2.5 for Item 7 (systematic methods used to search for evidence), between 1.5 and 2 for Item 8 (the criteria for selecting the evidence

are clearly described), between 2.5 and 4 for Item 9 (the strengths and limitations of the body of evidence are clearly described), and between 1.5 and 4.5 for Item 12 (explicit link between recommendations and supporting evidence). These items in the rigor of development domain highlighted the lack of transparent reports for the evidence basis and the link between evidence and the SOC8 recommendations. For the applicability domain, the median scores ranged between 3.0 and 4.0 for Item 19 (advice and/or tools for implementing recommendations) and between 1.0 and 1.5 for Item 21 (monitoring and/or auditing criteria). This suggests limited practicality of the recommendations. For all chapters, the median scores were 4 for Item 22 (the views of the funding body have not influenced the content of the guideline) and 3 to 3.5 for Item 23 (competing interests of guideline development group members have been recorded and addressed). The results on these two items suggested a potential lack of editorial independence.

Figure 1 shows the distribution of scores among assessors for the adolescent chapter. A similar trend exists across all six chapters.

Across six chapters, only two of eight assessors would recommend use without any modifications (Table 3). Six of eight assessors offered their reasoning for either recommending or not recommending use of the guidelines. The complete list of deidentified comments is available from the corresponding author. The main reason for recommending or recommending with modification is the guidelines' relevance to the needs of healthcare providers for trans-identified children and adolescents. Meanwhile, significant methodological concerns were cited as the main reason for not recommending the guideline chapters for use.

Table 1 AGREE II domain-specific scores and overall quality scores for six chapters

Domain	Adolescents	Children	Hormone therapy	Mental health	Primary care	Surgery
Scope and Purpose	61%	66%	69%	62%	60%	58%
Stakeholder Involvement	56%	56%	49%	56%	60%	56%
Rigor of Development	47%	40%	44%	39%	40%	44%
Clarity of Presentation	55%	56%	64%	52%	52%	59%
Applicability	28%	29%	40%	31%	33%	34%
Editorial Independence	43%	44%	43%	44%	43%	43%
Overall Quality (median (range))	3.5 (2 to 5)	4.0 (1 to 5)	4.0 (1 to 5)	3.5 (2 to 5)	3.5 (1 to 5)	3.5 (2 to 5)
Overall Quality (mean)	3.4	3.5	3.8	3.4	3.5	3.5

Table 2 Median and mean scores (on a scale of 1 to 7, the higher the better) assigned by assessors for each item

Domain	Item	Adolescents		Children		Hormone therapy		Mental health		Primary care		Surgery	
		Mean	Median (Range)	Mean	Median (Range)	Mean	Median (Range)	Mean	Median (Range)	Mean	Median (Range)	Mean	Median (Range)
Scope and purpose	1. The overall objective(s) of the guideline is (are) specifically described	5.1	5.5 (3-7)	5.0	5.5 (1-7)	5.3	6.0 (2-7)	5.1	5.5 (3-7)	5.0	5.0 (3-7)	5.0	5.0 (3-7)
	2. The health question(s) covered by the guideline is (are) specifically described	3.6	3.5 (1-7)	4.5	5.0 (1-7)	4.4	4.5 (1-7)	3.8	3.5 (1-7)	3.8	4.0 (1-7)	3.8	4.0 (1-7)
	3. The population (patients, public, etc.) to whom the guideline is meant to apply is specifically described	5.3	5.5 (3-7)	5.4	5.5 (1-7)	5.8	6.0 (2-7)	5.3	5.5 (3-7)	5.0	5.0 (3-7)	4.8	5.0 (1-7)
	4. The guideline development group includes individuals from all the relevant professional groups	4.1	5.0 (2-7)	3.9	5.0 (1-7)	3.6	4.5 (1-7)	3.9	4.5 (1-7)	4.4	5.0 (2-7)	4.1	5.0 (1-7)
	5. The views and preferences of the target population (patients, public, etc.) have been sought	3.9	3.5 (1-7)	3.4	3.0 (1-7)	3.4	3.0 (1-7)	4.3	4.5 (2-7)	3.9	3.0 (2-7)	4.0	4.0 (1-7)
	6. The target users of the guideline are clearly defined	5.1	5.0 (3-7)	5.9	6.0 (4-7)	4.8	5.5 (1-7)	4.9	5.0 (2-7)	5.6	6.0 (3-7)	5.0	5.5 (2-7)
Rigor of development	7. Systematic methods were used to search for evidence	2.9	2.5 (1-7)	2.5	1.5 (1-7)	2.9	2.0 (1-7)	3.0	2.5 (1-7)	3.1	2.5 (1-7)	2.8	2.0 (1-7)
	8. The criteria for selecting the evidence are clearly described	3.3	2.0 (1-7)	2.3	1.5 (1-5)	2.6	2.0 (1-7)	2.3	1.5 (1-5)	2.5	2.0 (1-7)	2.8	2.0 (1-7)
	9. The strengths and limitations of the body of evidence are clearly described	4.0	4.0 (1-7)	3.3	2.5 (1-7)	3.6	3.0 (1-7)	2.8	3.0 (1-5)	3.1	3.5 (1-5)	3.8	3.5 (1-7)
Clarity of presentation	10. The methods for formulating the recommendations are clearly described	4.3	4.5 (1-7)	3.9	3.5 (1-7)	3.9	3.0 (2-7)	3.6	3.0 (1-7)	3.3	2.5 (1-7)	3.8	3.0 (2-7)
	11. The health benefits, side effects and risks have been considered in formulating the recommendations	4.1	3.5 (2-7)	3.8	3.5 (1-7)	3.9	3.5 (1-7)	3.8	3.5 (1-7)	4.0	4.5 (1-7)	3.9	3.5 (2-7)
	12. There is an explicit link between the recommendations and the supporting evidence	2.9	1.5 (1-7)	3.3	2.0 (1-7)	3.8	4.5 (1-6)	3.5	3.0 (1-7)	3.3	3.0 (1-7)	4.1	4.5 (1-7)
	13. The guideline has been externally reviewed by experts prior to its publication	4.1	4.0 (2-7)	4.0	3.5 (2-7)	4.0	3.5 (2-7)	3.9	3.5 (1-7)	3.8	3.5 (1-7)	3.6	3.5 (1-7)
	14. A procedure for updating the guideline is provided	5.0	5.0 (3-7)	4.5	4.0 (2-7)	4.5	4.0 (2-7)	4.1	4.0 (1-7)	4.3	4.0 (1-7)	4.5	4.5 (3-7)
	15. The recommendations are specific and unambiguous	3.6	3.5 (1-7)	3.8	4.0 (1-6)	4.9	5.0 (2-7)	3.8	3.5 (2-7)	3.9	4.0 (1-7)	3.9	4.0 (1-7)
	16. The different options for management of the condition or health issue are clearly presented	3.9	3.5 (1-7)	3.8	3.5 (1-7)	4.0	4.5 (1-7)	3.0	3.0 (1-7)	3.1	2.0 (1-7)	3.9	3.5 (1-7)
	17. Key recommendations are easily identifiable	5.4	5.5 (3-7)	5.5	5.5 (4-7)	5.6	6.0 (3-7)	5.6	6.0 (3-7)	5.4	5.5 (3-7)	5.9	6.0 (4-7)

Table 2 (continued)

Domain	Item	Adolescents		Children		Hormone therapy		Mental health		Primary care		Surgery	
		Mean	Median (Range)	Mean	Median (Range)	Mean	Median (Range)	Mean	Median (Range)	Mean	Median (Range)	Mean	Median (Range)
Applicability	18. The guideline describes facilitators and barriers to its application	3.8	3.5 (1-7)	3.8	4.0 (1-7)	4.5	5.0 (1-7)	4.1	4.5 (1-7)	3.9	4.5 (1-7)	3.9	3.5 (1-7)
	19. The guideline provides advice and/or tools on how the recommendations can be put into practice	2.9	3.0 (1-4)	3.0	3.0 (1-5)	3.8	4.0 (1-6)	3.0	3.0 (1-5)	3.3	3.5 (1-5)	3.4	3.5 (1-6)
	20. The potential resource implications of applying the recommendations have been considered	2.5	1.0 (1-7)	2.4	1.5 (1-7)	3.1	2.0 (1-7)	2.6	2.0 (1-5)	3.0	2.0 (1-7)	3.0	2.0 (1-7)
Editorial independence	21. The guideline presents monitoring and/or auditing criteria	1.6	1.0 (1-4)	1.9	1.5 (1-3)	2.3	1.5 (1-6)	1.6	1.0 (1-4)	1.9	1.5 (1-4)	2.0	1.5 (1-4)
	22. The views of the funding body have not influenced the content of the guideline	3.6	4.0 (1-7)	3.8	4.0 (1-7)	3.5	4.0 (1-7)	3.6	4.0 (1-7)	3.6	4.0 (1-7)	3.5	4.0 (1-7)
	23. Competing interests of guideline development group members have been recorded and addressed	3.5	3.0 (1-7)	3.5	3.0 (1-7)	3.6	3.5 (1-7)	3.6	3.5 (1-7)	3.5	3.0 (1-7)	3.6	3.5 (1-7)

Discussion

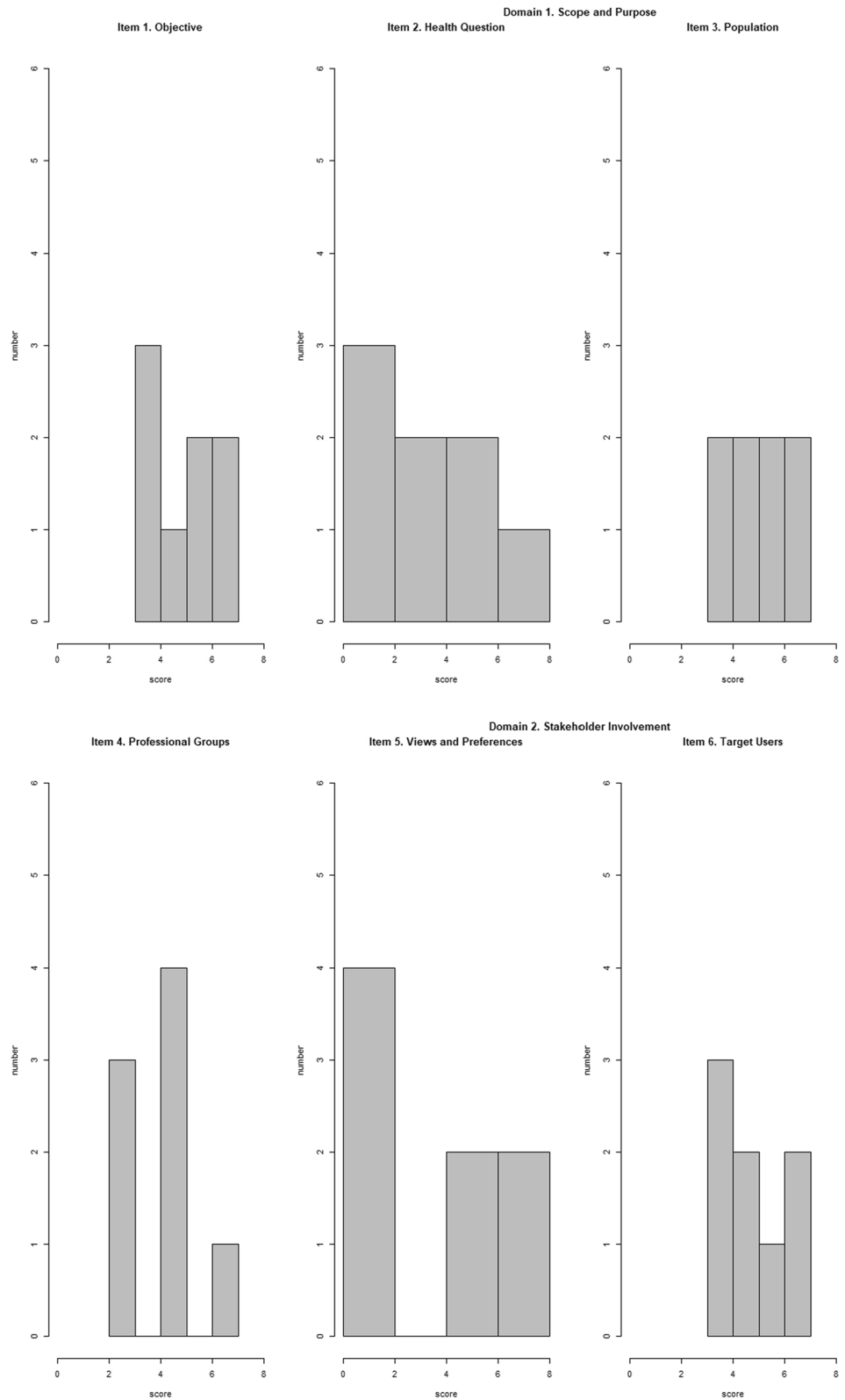
Main Findings

This study examined the guidelines' quality from the perspectives of healthcare providers for trans-identified children and adolescents and guideline methodologists. Our assessment identified significant shortcomings in WPATH's SOC8: lack of rigor of development, probably compromised editorial independence, and limited applicability. It is important to distinguish between the quality of the evidence base and the quality of guideline development methodology. While some clinical fields, including gender dysphoria in children and adolescents, may have limited or evolving evidence, this does not preclude the development of high-quality guidelines. The AGREE II assessment evaluates the rigor and transparency of the guideline development process, independent of the strength of the underlying evidence. SOC8 received low scores not due to the limitations of the evidence itself, but because of weaknesses in how the recommendations were formulated.

Regarding rigor of development, WPATH claimed the systematic reviews of evidence were conducted; however, our assessors identified issues with the methods and results of systematic reviews (Items 7, 8, and 9) and the process of incorporating the systematic reviews in the formulation of recommendations (Items 10, 11, and 12). These limitations led to concerns that the recommendations formulated may have diverged from those that could be justified considering the evidence basis and its certainty. In June 2024, news reports suggested that the systematic review team was not able to submit their systematic review manuscripts for peer-reviewed publication, claiming WPATH had "authority to influence the [systematic review] team's output" (The Economist, 2024). Our assessment using the AGREE II tool was not able to verify the reason, though it reflected the lack of transparency in the evidence reports. Furthermore, compromised editorial independence refers to a lack of transparency and safeguards in managing conflicts of interest, rather than merely the professional affiliations of the authors. While it is expected that experts in a field will be involved in guideline development, high-quality guidelines implement structured processes to ensure that conflicts—whether financial, intellectual, or professional—do not unduly shape recommendations. SOC8 does not clearly describe how conflicts of interest were identified and addressed, raising concerns about the extent to which editorial independence was maintained in the formulation of recommendations.

According to the independent review by NHS England (Taylor et al., 2024; The Cass Review, 2024), guidelines on transgender healthcare for minors lacked adherence to

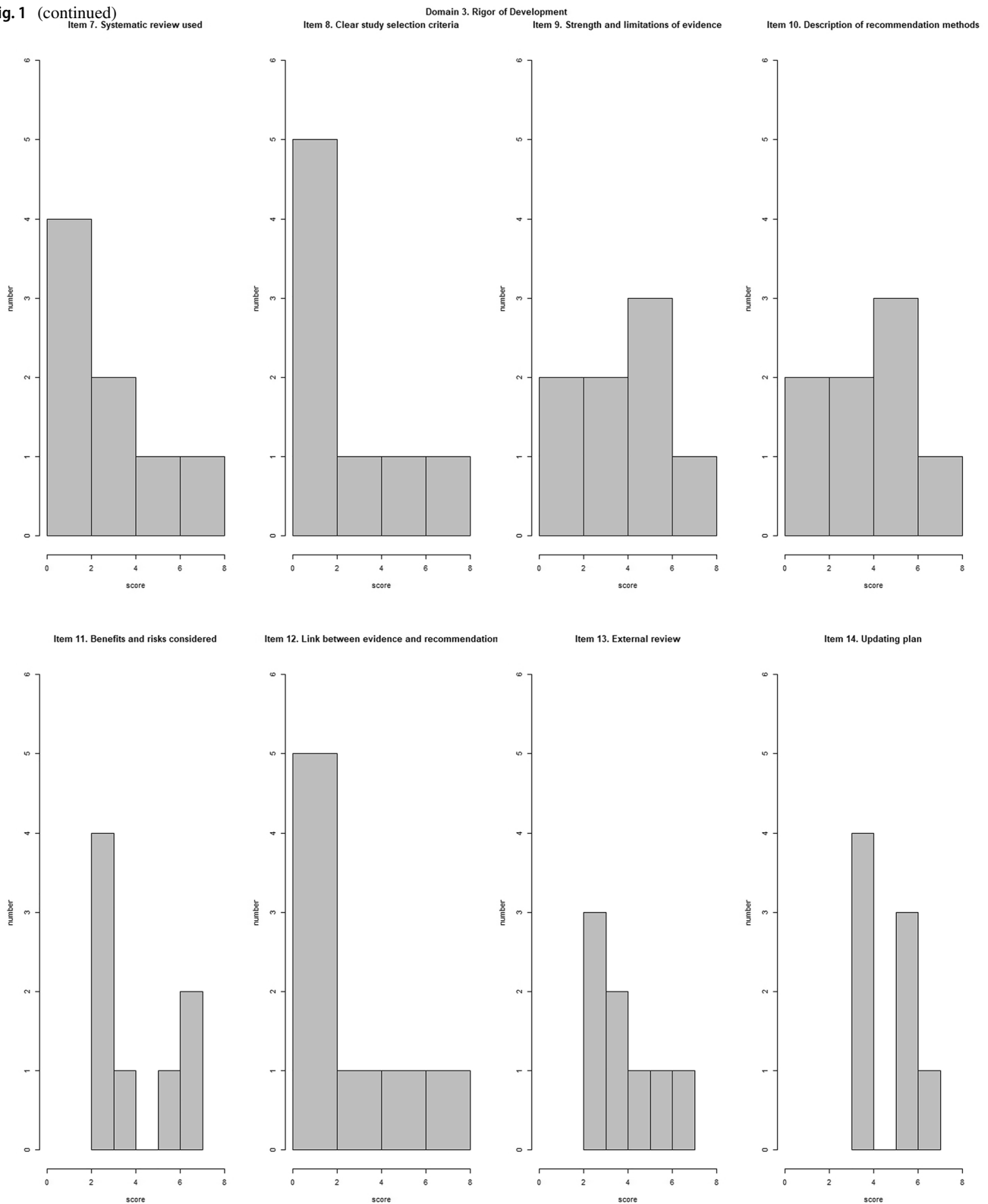
Fig. 1 Distribution of scores for all items for adolescent chapter. This figure shows the distribution of item scores from eight assessors for all 23 items, categorized by quality domains. Each bar represents the number of assessors assigning a certain score to this item.



international standards, particularly in rigor and transparency. Our domain-specific scores aligned with these findings. Similarly, a previous assessment of SOC7 guidelines

(Coleman et al., 2012) (an earlier version developed by WPATH) by Dahlen et al. identified limitations in rigor, clarity, applicability, and editorial independence (Dahlen

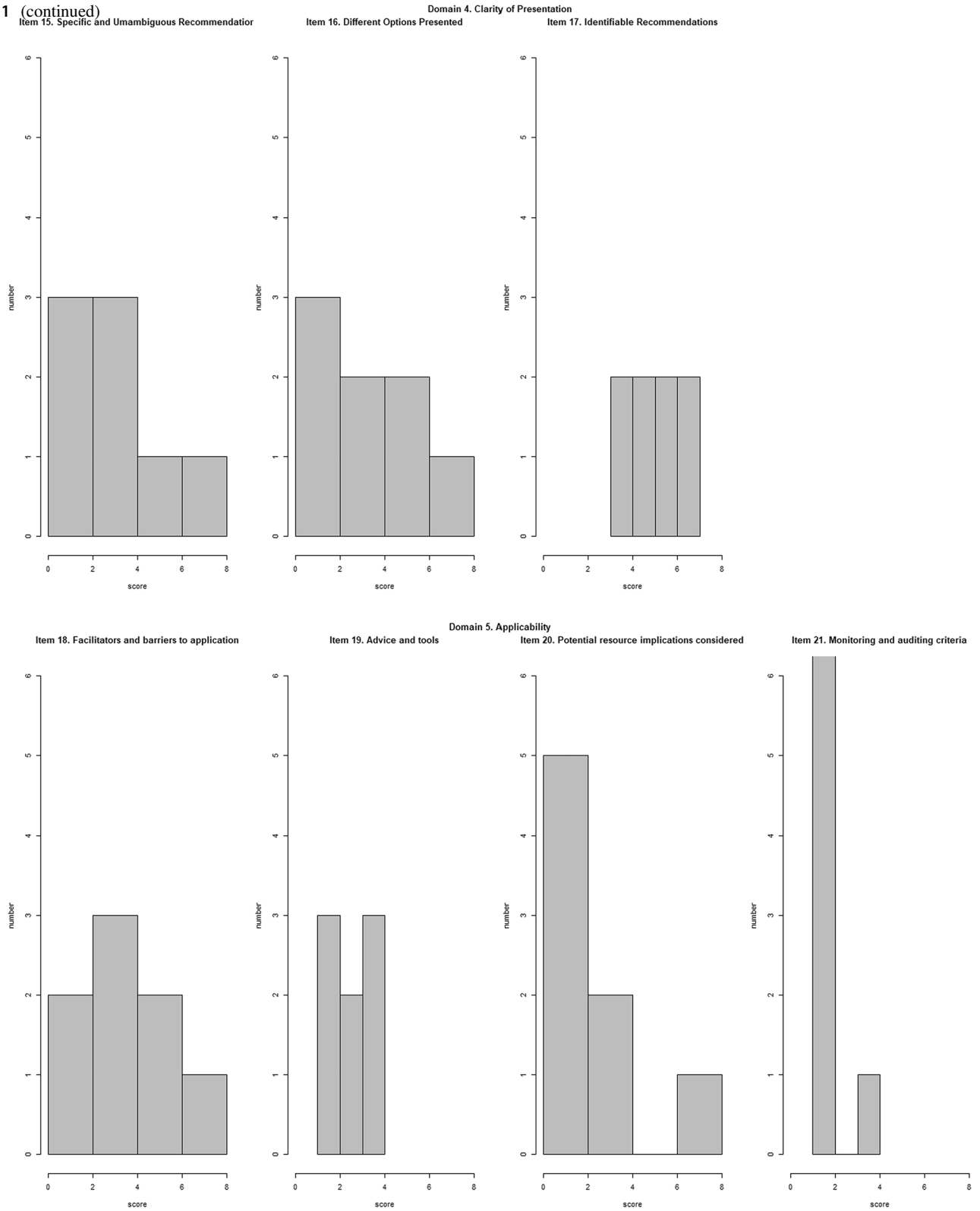
Fig. 1 (continued)



et al., 2021). Moreover, compared to AGREE II evaluations of clinical guidelines in other pediatric fields, SOC7

received lower scores, especially in domains related to systematic evidence review and conflict of interest

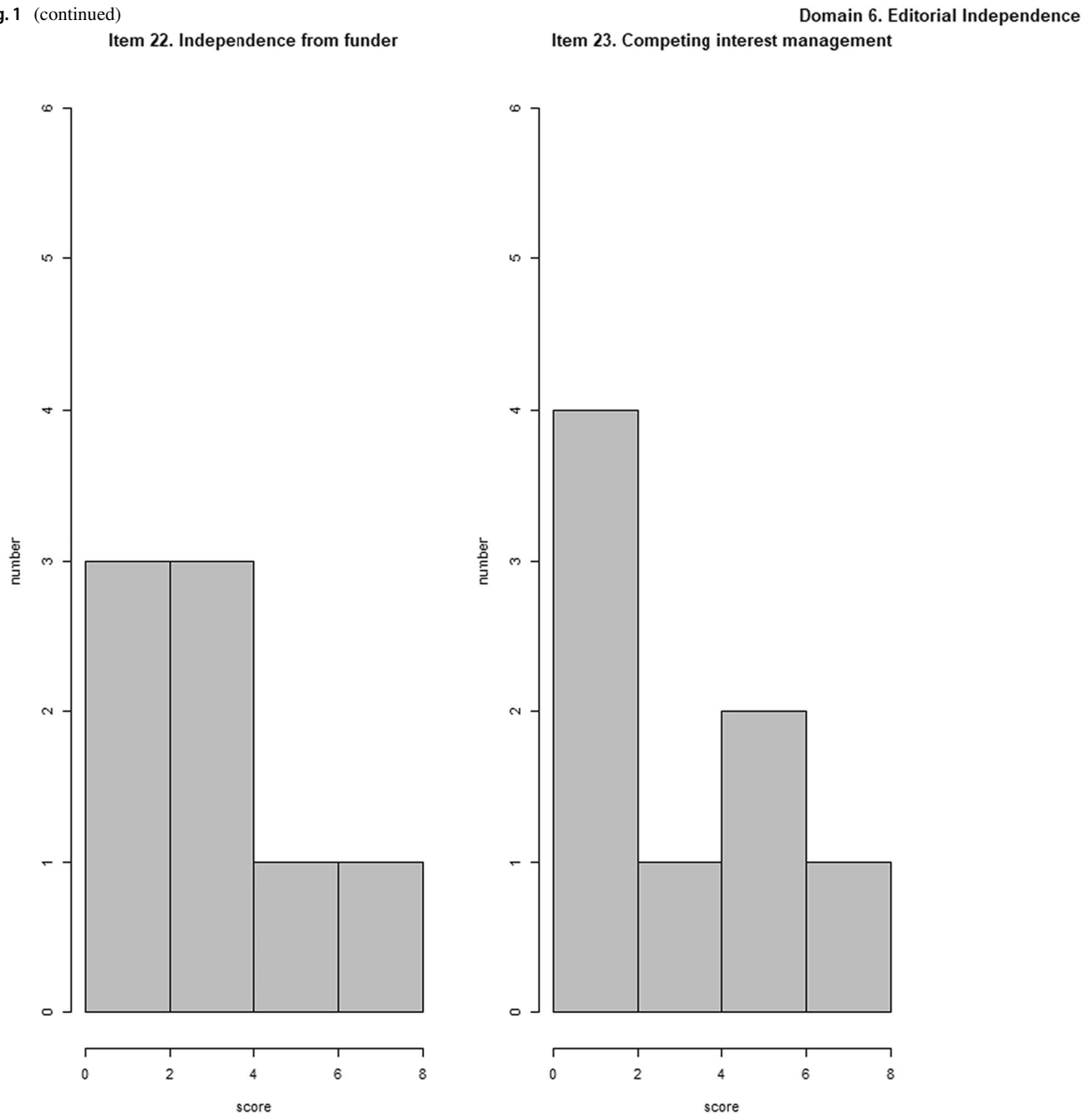
Fig. 1 (continued)



management (Dahlen et al., 2021). While SOC8 may show some improvements over SOC7, our evaluation indicates

that similar methodological weaknesses persist. Given its influence on clinical practice and policy, it is crucial that

Fig. 1 (continued)



SOC8 meets the same rigorous standards expected of other pediatric guidelines.

Implications

WPATH's SOC8, translated into many languages, stands as the most globally influential in this area (Human Rights Campaign Foundation, 2024; Restar et al., 2024; World Professional Association for Transgender Health, 2024). Yet, our analysis revealed important flaws in their development, notably a lack of scientific rigor and opaque conflict

of interest management, which could undermine the trust in these recommendations. Healthcare providers, associations, and policymakers should be cautious in uncritically adopting or endorsing SOC8. The methodology limitations of the guidelines could hinder the delivery of optimal healthcare to vulnerable youth.

Our study also found that the applicability of the guidelines to support clinical practice was limited; for instance, while the SOC8 advocates for a comprehensive assessment of transgendered adolescents, they fail to provide necessary resources, tools, or guidance on overcoming barriers to such

Table 3 Summary of recommendation on the use of guideline chapters

Recommendation	Adolescents (n)	Children (n)	Hormone therapy (n)	Mental health (n)	Primary care (n)	Surgery (n)	Quote from assessors
Recommending use of guideline	2	2	2	2	2	2	“This chapter is very well written and presents very important psychological approaches important in everyday clinic for me”
Recommending use of guideline with modification	3	3	4	4	3	4	“Recommendation with reservations because of the relatively weak evidence base, but I think we still need to have this guideline in order to be able provide some treatment to these [target users]” “The recommended comprehensive assessment is important and given the developments that professionals in some countries are threatened to be seen as attempting “conversion” if they perform any assessment, the guideline is an improvement. But on the other hand, it is erratic in ignoring that the evidence base for medical interventions is vague and therefore should not be used as a golden standard”
Recommending no use of guideline	3	3	2	2	3	2	“Major issue with reporting methods, results, and process of incorporation. Not clear nor transparent. Cannot endorse use of recommendations.”

assessments. The challenges in implementing, monitoring, and auditing SOC8 could potentially lead to unintended negative outcomes or inadequate care. Therefore, we call for concerted efforts to create high-quality, evidence-based guidelines that better serve the healthcare needs of trans-identified children and adolescents. Moreover, our findings underscore the importance of enhancing the applicability of guideline recommendations in future development and adoption processes.

Strengths and Limitations

Our study adeptly managed conflicts of interests and minimized strong opinions against or for gender-affirming care. We adopted a process of screening and minimizing conflict of interest of potential assessors and recruited a group of assessors who represent a broad spectrum of healthcare providers, focusing on objective evidence rather than personal experiences or agendas. These assessors were open to prescribing gender-affirming care when benefits outweigh harm and to discontinuing it when evidence indicates net harm. Despite the inherent challenges in assembling an entirely unbiased group, this group of assessors was open to different options as suggested by evidence and was not embedded with critical

attitudes toward gender-affirming care. They are likely representative of the broader health professional community committed to evidence-based medicine and the scientific rigor of guideline development. Another significant strength was our use of the AGREE II tool, which involved comprehensive training, calibration, and feedback from investigators, ensuring assessors were well-prepared to evaluate the SOC8. The application of the AGREE II instrument, a standardized tool designed to assess the methodological rigor of guideline development, is inherently independent of any clinical perspectives or clinical topics. Additionally, the study utilized a descriptive analytical approach, presenting individual ratings from assessors without modification or interpretation. No inferential statistical analysis or synthesis was performed beyond summarizing AGREE II domain scores and assessor responses. Each score reflects the independent judgment of the participating assessors, further reinforcing the transparency of our findings. Given the standardized nature of the AGREE II instrument, the rigorous conflict of interest screening process, and the descriptive presentation of results, the study findings were not systematically influenced by any particular perspective from the investigators or the funder.

Finally, we separately assessed six chapters relevant to healthcare for trans-identified children and adolescents; thus,

we were able to provide a granular level assessment of those guideline chapters.

Our study had several limitations. Firstly, although we employed attitude screening to exclude individuals with strong preconceived stances on gender-affirming care, verifying the authenticity of the stated attitudes was not feasible. Secondly, the sample size was small. Our sample size was larger than the recommendation by the AGREE II (2 to 4 assessors), and empirical studies (Dahlen et al., 2021; Leibowitz, 2023; Taylor et al., 2024). While our sample size was sufficient for assessing guideline quality based on AGREE II recommendations, it may be considered small in terms of achieving broader expertise diversity. Another limitation is our decision not to seek consensus among assessors or to exclude outlier assessments. Instead, we chose to present the assessors' evaluations independently, which may have resulted in the inclusion of lenient assessments that overlooked the guidelines' evident shortcomings. Additionally, our assessment focused on the SOC8, rather than a broader scope of all clinical guidelines for gender dysphoria or incongruence. Thus, we were not able to examine the link between guidelines "through co-sponsorship" or circulatory citations (Taylor et al., 2024).

Conclusion

Our assessment revealed that WPATH's SOC8 guidelines have limitations in scientific and methodological rigor, applicability, and transparency in managing competing interests. Evidence-based guidelines addressing the needs of transgender children and adolescents are urgently needed, but the uncritical adoption or endorsement of WPATH's guidelines may result in a disservice or even harm to this vulnerable population. It is imperative that healthcare providers, researchers, and policymakers recognize and address the limitations of WPATH's SOC8.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s10508-025-03399-6>.

Acknowledgements We thank all the assessors for their contributions to this project. The authors would like to acknowledge researchers who assisted in study design and participant recruitment during the early phase of this project.

Author Contributions Participant recruitment and data collection were involved by YZ, DJ, RK, MK, JP, and JJYN; Analysis and interpretation of results and draft manuscript preparation were performed by YZ; all authors reviewed and edited the manuscript and approved the final version of the manuscript.

Funding This study was funded by the Society for Evidence-Based Gender Medicine (SEGM). The funder had no role in the study design, data collection, data analysis, interpretation, and reporting of the results, or drafting of the manuscript.

Data Availability The primary data that support the findings of this study are available on request from the corresponding author. The data are not publicly available due to privacy or ethical restrictions.

Declarations

Conflict of interest ZY provided consultation to SEGM. The assessors (JD, KR, KM, PJ, and YNJJ) received reimbursement of \$1,500 USD for their estimated 30 to 40 h of contribution from the research project. The authors declare no other conflicts of interest.

Ethics Approval Due to the focus on guideline quality appraisal by health professionals, this project received a waiver from the Hamilton Integrated Research Ethics Board (HiREB), in accordance with the Tri-Council Policy Statement: Ethical Conduct for Research Involving Humans (TCPS 2, Article 2.5) (Canadian Institutes of Health Research, 2018).

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

References

- Alonso-Coello, P., Oxman, A. D., Moberg, J., Brignardello-Petersen, R., Akl, E. A., Davoli, M., Treweek, S., Mustafa, R. A., Vandvik, P. O., Meerpohl, J., Guyatt, G. H., & Schünemann, H. J. (2016a). GRADE evidence to decision (EtD) frameworks: A systematic and transparent approach to making well informed healthcare choices. 2: Clinical practice guidelines. *British Medical Journal*, 353, Article i2089. <https://doi.org/10.1136/bmj.i2089>
- Alonso-Coello, P., Schünemann, H. J., Moberg, J., Brignardello-Petersen, R., Akl, E. A., Davoli, M., Treweek, S., Mustafa, R. A., Rada, G., Rosenbaum, S., Morelli, A., Guyatt, G. H., & Oxman, A. D. (2016b). GRADE evidence to decision (EtD) frameworks: A systematic and transparent approach to making well informed healthcare choices. 1: Introduction. *British Medical Journal*, 353, Article i2016. <https://doi.org/10.1136/bmj.i2016>
- Bahtsevani, C., Udén, G., & Willman, A. (2004). Outcomes of evidence-based clinical practice guidelines: A systematic review. *International Journal of Technology Assessment in Health Care*, 20(4), 427–433. <https://doi.org/10.1017/s026646230400131x>
- Brouwers, M., Kho, M., Browman, G., Burgers, J., Cluzeau, F., Feder, G., Fervers, B., Graham, I., Grimshaw, J., Hanna, S., Littlejohns, P., Makarski, J., & Zitzelsberger, L. (2010a). AGREE II: Advancing guideline development, reporting and evaluation in health care. *Canadian Medical Association Journal*, 182(18), E839–E842. <https://doi.org/10.1503/cmaj.090449>

- Brouwers, M., Kho, M., Browman, G., Burgers, J., Cluzeau, F., Feder, G., Fervers, B., Graham, I., Hanna, S., & Makarski, J. (2010b). Development of the AGREE II, part 2: Assessment of validity of items and tools to support application. *Canadian Medical Association Journal*, 182(10), E472–E478. <https://doi.org/10.1503/cmaj.091716>
- Canadian Institutes of Health Research, Natural Sciences and Engineering Research Council of Canada, & Social Sciences and Humanities Research Council of Canada. (2018). *Tri-Council policy statement: Ethical conduct for research involving humans*. https://ethics.gc.ca/eng/policy-politique_tcps2-eptc2_2018.html
- Coleman, E., Bockting, W., Botzer, M., Cohen-Kettenis, P., DeCuypere, G., Feldman, J., Fraser, L., Green, J., Knudson, G., Meyer, W. J., Monstrey, S., Adler, R. K., Brown, G. R., Devor, A. H., Ehrbar, R., Ettner, R., Eyler, E., Garofalo, R., Karasic, D. H., ... Zucker, K. (2012). Standards of care for the health of transsexual, transgender, and gender-nonconforming people, Version 7. *International Journal of Transgenderism*, 13(4), 165–232. <https://doi.org/10.1080/15532739.2011.700873>
- Coleman, E., Radix, A. E., Bouman, W. P., Brown, G. R., de Vries, A. L. C., Deutsch, M. B., Ettner, R., Fraser, L., Goodman, M., Green, J., Hancock, A. B., Johnson, T. W., Karasic, D. H., Knudson, G. A., Leibowitz, S. F., Meyer-Bahlburg, H. F. L., Monstrey, S. J., Motmans, J., Nahata, L., ... Arcelus, J. (2022). Standards of care for the health of transgender and gender diverse people, Version 8. *International Journal of Transgender Health*, 23(Suppl 1), S1–S259. <https://doi.org/10.1080/26895269.2022.2100644>
- Dahlen, S., Connolly, D., Arif, I., Junejo, M. H., Bewley, S., & Meads, C. (2021). International clinical practice guidelines for gender minority/trans people: Systematic review and quality assessment. *British Medical Journal Open*, 11(4), Article e048943. <https://doi.org/10.1136/bmjopen-2021-048943>
- Guyatt, G. H., Akl, E. A., Hirsh, J., Kearon, C., Crowther, M., Gutterman, D., Lewis, S. Z., Nathanson, I., Jaeschke, R., & Schünemann, H. (2010). The vexing problem of guidelines and conflict of interest: A potential solution. *Annals of Internal Medicine*, 152(11), 738–741. <https://doi.org/10.7326/0003-4819-152-11-201006010-00254>
- Guyatt, G. H., Oxman, A. D., Kunz, R., Falck-Ytter, Y., Vist, G. E., Liberati, A., & Schünemann, H. J. (2008). Going from evidence to recommendations. *British Medical Journal*, 336(7652), 1049–1051. <https://doi.org/10.1136/bmj.39493.646875.AE>
- Hansen, C., Lundh, A., Rasmussen, K., & Hróbjartsson, A. (2019). Financial conflicts of interest in systematic reviews: Associations with results, conclusions, and methodological quality. *Cochrane Database of Systematic Reviews*, 8(8), Article Mr000047. <https://doi.org/10.1002/14651858.MR000047.pub2>
- Haynes, R. B., Devereaux, P. J., & Guyatt, G. H. (2002). Clinical expertise in the era of evidence-based medicine and patient choice. *Evidence-Based Medicine*, 7(2), 36–38.
- Human Rights Campaign Foundation. (2024). *Supporting & caring for transgender children*. <https://www.hrc.org/resources/supporting-caring-for-transgender-children>
- Institute of Medicine Committee on Standards for Developing Trustworthy Clinical Practice Guidelines. (2011). *Clinical practice guidelines we can trust*. National Academies Press. <https://doi.org/10.17226/13058>
- Leibowitz, S. F. (2023). Assessment of transgender and gender-diverse adolescents: Incorporating the World Professional Association of Transgender Health Standard of Care 8th Edition. *Child and Adolescent Psychiatric Clinics of North America*, 32(4), 707–718. <https://doi.org/10.1016/j.chc.2023.05.009>
- Restar, A. J., Layland, E. K., Davis, B., Thompson, H., & Streed, C. (2024). The public health crisis state of transgender health care and policy. *American Journal of Public Health*, 114(2), 161–163. <https://doi.org/10.2105/ajph.2023.307523>
- Rosenfeld, R. M., Shiffman, R. N., & Robertson, P. (2013). Clinical practice guideline development manual, third edition: A quality-driven approach for translating evidence into action. *Otolaryngology-Head and Neck Surgery*, 148(1 Suppl), S1–55. <https://doi.org/10.1177/0194599812467004>
- RStudio Team. (2024). *RStudio: Integrated development for R* [Computer software]. <http://www.rstudio.com/>
- Sackett, D. L., Rosenberg, W. M., Gray, J. A., Haynes, R. B., & Richardson, W. S. (1996). Evidence based medicine: What it is and what it isn't. *British Medical Journal*, 312(7023), 71–72.
- Shiffman, R. N., Marcuse, E. K., Moyer, V. A., Neuspiel, D. R., Hodgson, E. S., Glade, G., Harbaugh, N., Jr., Miller, M. R., Sevilla, X., Simpson, L., & Takata, G. (2008). Toward transparent clinical policies. *Pediatrics*, 121(3), 643–646. <https://doi.org/10.1542/peds.2007-3624>
- Siemieniuk, A. R., Agoritsas, T., Macdonald, H., Guyatt, H. G., Brandt, L., & Vandvik, P. (2016). Introduction to BMJ rapid recommendations. *British Medical Journal*, 354, Article i5191. <https://doi.org/10.1136/bmj.i5191>
- Stone, E. G. (2018). Evidence-based medicine and bioethics: Implications for health care organizations, clinicians, and patients. *The Permanente Journal*, 22, 18–030. <https://doi.org/10.7812/tpj/18-030>
- Taylor, J., Hall, R., Heathcote, C., Hewitt, C. E., Langton, T., & Fraser, L. (2024). Clinical guidelines for children and adolescents experiencing gender dysphoria or incongruence: A systematic review of recommendations (part 2). *Archives of Disease in Childhood*, 109, s73–s82. <https://doi.org/10.1136/archdischild-2023-326500>
- The Cass Review. (2024). *Independent review of gender identity services for children and young people*. <https://cass.independent-review.uk/>
- The Economist. (2024, June 27). Research into trans medicine has been manipulated. *The Economist*. <https://www.economist.com/united-states/2024/06/27/research-into-trans-medicine-has-been-manipulated>
- Varkey, B. (2021). Principles of clinical ethics and their application to practice. *Medical Principles and Practice*, 30(1), 17–28. <https://doi.org/10.1159/000509119>
- Woolf, S. H., Grol, R., Hutchinson, A., Eccles, M., & Grimshaw, J. (1999). Clinical guidelines: Potential benefits, limitations, and harms of clinical guidelines. *British Medical Journal*, 318(7182), 527–530. <https://doi.org/10.1136/bmj.318.7182.527>
- World Professional Association for Transgender Health. (2024). *Standards of Care Version 8*. <https://www.wpath.org/soc8>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.