

Sélection des résultats et biais d'interprétation dans l'étude de Chen et al., 2023 : Analyse critique d'une étude phare sur le traitement hormonal chez les jeunes s'identifiant transgenres

Introduction

L'étude de Chen et al. (2023), publiée dans le *New England Journal of Medicine*, a suivi pendant deux ans l'évolution psychosociale de 315 adolescents transgenres et non binaires traités par hormones sexuelles croisées (*Gender Affirming Hormones*). Largement saluée dans les milieux professionnels et médiatiques, cette publication est souvent présentée comme une preuve de l'efficacité de ces interventions médicales chez les mineurs. Cette publication est souvent présentée comme une preuve de l'efficacité de ces interventions médicales chez les mineurs. McNamara et al. (2024) ont ainsi reproché aux [revues systématiques](#) de l'Université de York, réalisées pour le Cass Review (2024), d'avoir ignoré certaines « études à fort impact » censées démontrer cette efficacité -au premier rang desquelles l'étude de Chen et al. (2023).

Une analyse approfondie du protocole enregistré et des résultats publiés révèle toutefois de sérieuses limites méthodologiques et interprétatives. Ce texte propose une relecture critique de l'étude, en montrant que les auteurs :

1. ont profondément modifié les hypothèses préenregistrées et écarté, sans justification, plusieurs variables psychosociales clés pourtant mesurées, tout en affirmant garantir la fidélité de l'étude au protocole : une contradiction manifeste qui soulève de graves questions de transparence scientifique ;
2. ont introduit *a posteriori* une variable secondaire, la congruence de l'apparence, élevée au rang de résultat principal ;
3. ont sélectionné les résultats les plus favorables en passant sous silence ceux allant à l'encontre de leur hypothèse pré-enregistrée ;
4. ont interprété de manière causale des associations non démontrées, sans groupe témoin ni prise en compte des traitements concomitants ;
5. ont présenté des conclusions trompeuses, en contradiction avec la faiblesse des données rapportées.

En restituant les données dans leur contexte et en analysant les écarts méthodologiques, cette critique souligne les risques d'une interprétation biaisée dans un domaine médical aux enjeux cliniques, éthiques et politiques considérables.

1. Sélection post hoc des variables : une discordance avec le protocole enregistré

L'article de Chen et al. (2023) se présente comme une évaluation des effets des hormones sexuelles croisées (GAH) sur le bien-être psychosocial des jeunes transgenres. Cependant, comme l'a longuement décrit Singal (2023), les hypothèses formulées et les variables rapportées dans la publication finale (Chen et al. 2023, p. 241) s'écartent notablement de celles du protocole enregistré (Chen et al., 2021, p. 36/44).

Le tableau 1 compare les formulations d'hypothèses figurant dans les deux versions :

Tableau 1. Hypothèses du protocole (Chen et al., 2021) vs. hypothèses rapportées dans l'article (Chen et al., 2023)

Hypothèses définies dans le protocole enregistré (Chen et al. 2021, p. 36/44)	Hypothèses dans l'article final (Chen et al. 2023, p. 241)
<i>Les patients traités par hormones affirmant le genre présenteront une diminution des symptômes d'anxiété et de dépression, de dysphorie de genre, d'automutilation, de symptômes liés aux traumatismes et de suicidalité, ainsi qu'une augmentation de l'estime corporelle et de la qualité de vie au cours du temps.</i>	<i>Nous avons émis l'hypothèse que la congruence de l'apparence, l'affect positif et la satisfaction de vie augmenteraient, et que les symptômes de dépression et d'anxiété diminueraient.</i> <i>Nous avons également formulé l'hypothèse que ces améliorations seraient secondaires au traitement de la dysphorie de genre, de sorte qu'une augmentation de la congruence de l'apparence serait associée à des améliorations concomitantes des résultats psychosociaux.</i>

La modification des hypothèses du protocole de l'étude a eu les conséquences suivantes :

- **75 % des variables** clés (6 sur 8) dont les chercheurs avaient émis l'hypothèse qu'elles s'amélioreraient sous l'effet des hormones, ont **été mises à l'écart**, sans justification : Suicidalité, Dysphorie de genre, Qualité de vie, Estime corporelle, Troubles liés aux traumatismes, Automutilation. Elles ont été pourtant mesurées durant cette étude (Chen et al., 2021, p. 22 à 28/44).
- Trois nouvelles variables sont rajoutées sans justification, **dont la congruence de l'apparence**, qui n'est pas classée comme un résultat psychosocial par les auteurs.

Le tableau 2 illustre l'écart entre les variables mentionnées dans les hypothèses du protocole enregistré et celles effectivement rapportées dans la publication finale de l'étude.

Tableau 2. Variables formulées dans les hypothèses du protocole de l'étude (Chen et al., 2021) et variables rapportées dans la publication finale de Chen et al. (2023)

Variables dont les chercheurs avaient prédit une amélioration (Chen et al., 2021, p. 36/44)	Variables mesurées durant l'étude (Chen et al., 2021, p. 22-28/44)	Variables rapportées dans la publication finale (Chen et al., 2023)
Anxiété	Oui	Oui
Dépression	Oui	Oui
Suicidalité	Oui	Non
Dysphorie de genre	Oui	Non
Qualité de vie	Oui	Non
Estime corporelle	Oui	Non
Troubles liés aux traumatismes	Oui	Non
Automutilation	Oui	Non

Tableau élaboré par l'auteure à partir de l'analyse du protocole enregistré (Chen et al., 2021) et de la publication finale (Chen et al., 2023)

Ce type de glissement entre hypothèses protocolaires et publication finale constitue un exemple classique de *HARKing* (*Hypothesizing After the Results are Known*) (Kerr, 1998). Il s'agit d'une reformulation post hoc des hypothèses en fonction des résultats obtenus, sans transparence sur les changements opérés.

De telles pratiques compromettent la rigueur scientifique, car elles biaisent l'interprétation des résultats en privilégiant les variables les plus susceptibles de confirmer une hypothèse favorable.

Elles rendent également profondément trompeuses la déclaration des auteurs selon laquelle : « *Les auteurs garantissent l'exactitude et l'exhaustivité des données, ainsi que la conformité de l'étude au protocole* ». (Chen et al., 2023, p. 247).

Une telle affirmation, en contradiction manifeste avec le déroulé réel de l'étude, est de nature à induire le lecteur en erreur sur la transparence et la validité méthodologique du travail présenté.

2. Surexploitation de la congruence d'apparence : un biais central

2.1. Une variable introduite a posteriori, absente du protocole et exploitée de manière sélective

La congruence de l'apparence est décrite par les auteurs comme « *le degré auquel les jeunes ressentent une adéquation entre leur genre et leur apparence physique* » (Chen et al., 2023, p. 241).

Son amélioration est décrite par les auteurs comme « *un objectif principal des hormones d'affirmation de genre* » (p. 245). Si tel est le cas, il est surprenant qu'elle ne soit pas mentionnée une seule fois dans le protocole enregistré en 2021.

Son introduction tardive, sans justification, constitue un biais méthodologique majeur.

Par ailleurs, cette variable correspond à l'une des deux sous-échelles de la *Transgender Congruence Scale* (TCS), une échelle multidimensionnelle administrée (Chen et al., 2023, p. 242).

Pourtant, **les auteurs ne rapportent ni ne discutent les résultats de l'autre sous-échelle**, qui porte sur l'acceptation de l'identité de genre. Cette omission, non justifiée, constitue un biais méthodologique manifeste (*cherry-picking*), car elle repose sur la sélection **a posteriori** de la composante la plus favorable aux hypothèses de l'étude.

Ce choix est d'autant plus contestable que dans une publication antérieure, les auteurs eux-mêmes qualifiaient la *Transgender Congruence Scale*, aux côtés de la *Gender Minority Stress and Resilience Scale*, comme l'un des « *meilleurs outils disponibles pour évaluer les facteurs proximaux et distaux contribuant à la dysphorie de genre* » (Olson-Kennedy et al., 2019). Son usage partiel ici, sans justification, suggère une sélection orientée des données au service d'une interprétation favorable.

2.2. Une mesure redondante et conceptuellement pauvre

La congruence de l'apparence est évaluée à l'aide de neuf affirmations telles que « *Mon corps physique reflète mon identité de genre* » (Huit et al., 2021), cotées sur une échelle de Likert en 5 points. Cette sous-échelle se concentre presque exclusivement sur la correspondance perçue

entre l'apparence corporelle et l'identité de genre, ce qui la rend particulièrement sensible aux effets attendus des hormones sexuelles croisées.

Sans surprise, cette variable est la seule à présenter une taille d'effet importante après 24 mois (Cohen's $d = -1,12$; tableau S5, [Supplementary Appendix](#)), aussi bien chez les filles que chez les garçons à la naissance.

Cependant, l'analyse du contenu des items (tableau 3) révèle une pauvreté conceptuelle marquée :

Tableau 3. Thème central de chacun des items de la sous-échelle Congruence de l'apparence fournie dans Kozee et al., 2012 ; Huit et al., 2021

Item	Thème central
1. Mon apparence extérieure reflétait mon identité de genre.	Apparence = identité
2. Je ressentais une unité entre mon identité de genre et mon corps.	Unité corps/identité
3. Mon apparence physique exprimait de manière adéquate mon identité de genre.	Apparence = identité
4. J'étais généralement à l'aise avec la manière dont les autres percevaient mon identité de genre en me regardant.	Perception d'autrui
5. Mon corps physique reflétait mon identité de genre.	Corps = identité
6. L'apparence actuelle de mon corps ne reflétait pas mon identité de genre. (<i>inversé</i>)	Corps $\neq$ identité
7. J'étais satisfait(e) de la manière dont mon apparence exprimait mon identité de genre.	Satisfaction apparence = identité
8. Je ne sentais pas que mon apparence reflétait mon identité de genre. (<i>inversé</i>)	Apparence $\neq$ identité
9. Je sentais que mon esprit et mon corps étaient en cohérence.	Cohérence interne

On constate que sept items sur neuf (1, 2, 3, 5, 6, 7, 8, soit environ **80 %**) évaluent exclusivement, sous des formulations légèrement différentes, la correspondance perçue entre le corps physique et l'identité de genre. Cette **redondance** tend à renforcer artificiellement la cohérence interne de l'échelle tout en limitant sévèrement sa capacité à capturer la complexité de l'expérience vécue.

L'interprétation du score de « congruence de l'apparence » comme indicateur majeur d'ajustement psychosocial doit donc être considérée avec prudence.

2.3. Une interprétation causale infondée

Les auteurs utilisent des modèles de croissance latente parallèles pour suggérer que l'amélioration de la congruence de l'apparence serait un mécanisme par lequel le traitement hormonal améliorerait le bien-être psychosocial. Or, cette interprétation repose uniquement sur des corrélations, sans analyse de médiation ni test d'hypothèse alternative (comme l'amélioration du bien-être influençant le sentiment de congruence).

Surtout, **les résultats eux-mêmes contredisent l'hypothèse avancée** : chez les garçons à la naissance, la congruence s'est améliorée de la même manière que chez les filles, mais, comme nous le verrons à la section suivante, **les résultats psychosociaux n'ont pas progressé**.

Ce paradoxe remet en cause l'idée que la congruence constitue un facteur causal central. Il suggère qu'un ou plusieurs facteurs non modélisés influencent les trajectoires observées, et que l'effet prêté à cette variable relève davantage d'un biais interprétatif que d'un lien démontré.

3. Déclarations des auteurs : un résumé trompeur

Dans leur article, Chen et al. (2023) avancent de nombreuses affirmations concernant l'évolution positive du fonctionnement psychosocial des jeunes suivis, l'efficacité du traitement hormonal affirmatif, et l'association entre la congruence d'apparence et les variables psychosociales.

Ces conclusions, largement reprises dans le résumé, les résultats et la discussion, méritent toutefois d'être examinées de manière critique à la lumière des résultats effectivement rapportés.

3.1. Résultats psychosociaux effectivement rapportés

3.1.1. Résultats psychosociaux chez les garçons à la naissance

Les auteurs reconnaissent eux-mêmes l'absence d'amélioration significative pour **trois des quatre variables** psychosociales rapportées : « *Les scores de dépression et d'anxiété ont diminué chez les jeunes de sexe féminin mais pas chez ceux de sexe masculin. De même, les scores de satisfaction de vie ont augmenté chez les jeunes de sexe féminin, mais pas chez ceux de sexe masculin (Fig. S3)* » (p. 244-245).

Concernant **l'affect positif**, les résultats sont similaires : aucune amélioration n'est observée, comme l'illustre la Figure S3B de l'annexe supplémentaire ([Supplementary Appendix](#)).

Ainsi, pour les garçons à la naissance, aucune amélioration significative n'est rapportée pour les variables psychosociales rapportées.

3.1.2. Résultats psychosociaux chez les filles à la naissance

Les auteurs de l'étude rapportent des améliorations statistiquement significatives des scores de dépression, d'anxiété et de satisfaction de vie au bout de 24 mois de traitement hormonal.

Toutefois, **les tailles d'effet demeurent modestes** (Cohen's $d \approx 0,20$; $0,25$; $0,39$) et **les variations cliniques, modestes** (Tableau S5, [Supplementary Appendix](#)).

Concernant la dépression, mesurée par l'échelle BDI-II (score de 0 à 63), le score moyen passe d'un peu plus de 15 à environ 12–13 après 24 mois (Figure S3D, [Supplementary Appendix](#)). Ce changement correspond à une évolution **mineure au sein de la même catégorie clinique**, celle de la « *perturbation légère de l'humeur* », et demeure **cliniquement insignifiant**.

L'évolution des autres variables psychosociales, notamment l'affect positif et la satisfaction de vie, suit une tendance similaire : **des effets positifs sont observés mais restent limités en ampleur clinique**.

3.1.3. Récapitulatif synthétique

Afin de synthétiser l'ensemble des résultats rapportés, ainsi que ceux évalués mais non publiés, le tableau 4 présente l'évolution des principales variables psychosociales sur la période de 24 mois suivant l'initiation du traitement hormonal sexuel croisé.

Ce tableau distingue les variables effectivement rapportées dans la publication finale de celles prévues dans le protocole mais absentes des résultats, afin de donner une vision d'ensemble fidèle à la conception initiale de l'étude.

Tableau 4 - Évolution des variables principales mesurées par Chen et al. [2023](#) sur une période de 24 mois suivant l'initiation d'un traitement hormonal sexuel croisé

Variable	Changement moyen (modèle conditionnel)	Taille d'effet (Cohen's d) tableau S5)	Amélioration significative/sexe
Congruence de l'apparence	+0,77/5	-1,12	Hommes et femmes
Affect positif	+2,36/100	-0,06	Seulement les femmes natales
Satisfaction de vie	+4,48/100	-0,39	Seulement les femmes natales
Dépression	-2,00/63	0,20	Seulement les femmes natales
Anxiété	-1,90/100	0,25	Seulement les femmes natales
Suicidalité	?	?	?
Dysphorie de genre	?	?	?
Qualité de vie	?	?	?
Estime corporelle	?	?	?
Troubles liés aux traumatismes	?	?	?
Automutilation	?	?	?
Acceptation de l'identité de genre*	?	?	?

Taille d'effet : 0,20 : *faible*, 0,50 : *modéré*, > 80 : *élevé*

Les changements moyens sont issus des modèles conditionnels ajustés pour les covariables sociodémographiques reportés dans le tableau S10 de l'[annexe supplémentaire](#). Les tailles d'effet sont celles rapportées dans le tableau S5.

« ? » indique les variables qui ont été mesurées mais non rapportées dans la publication finale, malgré leur présence dans les hypothèses du protocole enregistré

* Cette variable correspond à la deuxième sous-échelle de la Transgender Congruence Scale, mais n'a pas été rapportée.

Tableau élaboré par l'auteure à des fins de synthèse critique.

Les auteurs mentionnent par ailleurs **deux décès par suicide sur 315, au cours de la 1^{re} année** (soit un taux de 0,63 %), sans plus de détails. Ce taux, sans permettre d'inférer une causalité avec le traitement hormonal, apparaît **particulièrement élevé** si l'on considère, avec toutes les précautions nécessaires, le **taux cumulé de 0,3 %** observé dans le groupe orienté vers les services de genre dans l'étude de Ruuska et al. ([2024](#)), malgré un suivi médian nettement plus long (environ sept ans).

Dans une lettre à l'éditeur consacrée à cette étude, Biggs (2023) relève que le taux annuel de suicide, qu'il estime à 317 pour 100 000, est environ 24 fois supérieur à celui observé chez les 15 000 adolescents orientés vers la clinique de genre du Royaume-Uni entre 2010 et 2020, où il s'élevait à 13 pour 100 000 (Biggs, 2022).

Ce constat soulève d'autant plus de questions que le protocole de l'étude (Chen et al., 2021, p. 19/44) stipulait explicitement que les jeunes présentant des troubles psychiatriques graves, y compris une suicidalité manifeste, devaient être exclus avant l'inclusion : « *Critères d'exclusion (...) Manifestement en détresse (par ex. : suicidaire, homicide, comportement violent) au moment du recueil du consentement ou de l'évaluation de départ.* ».

3.2. Analyse des affirmations des auteurs au vu de ces résultats

Afin d'évaluer la fidélité des affirmations principales des auteurs à leurs propres données, le tableau 5 confronte les déclarations de Chen et al. (2023) aux résultats effectivement rapportés.

Tableau 5. Comparaison entre les affirmations des auteurs et les résultats observés

Affirmations des auteurs	Analyse critique
<i>Nos résultats ont montré une amélioration du fonctionnement psychosocial sur une période de deux ans de traitement hormonal affirmant le genre (GAH). (p. 249)</i>	Il s'agit uniquement des résultats psychosociaux sélectionnés et rapportés par les auteurs, à la suite d'une modification des hypothèses initialement formulées dans le protocole enregistré.
<i>Au cours de la période d'étude, la congruence de l'apparence, l'affect positif et la satisfaction de vie ont augmenté, tandis que les symptômes de dépression et d'anxiété ont diminué. (p. 241)</i> <i>Dans notre cohorte américaine de jeunes transgenres et non binaires traités par GAH, nous avons observé une diminution des symptômes dépressifs et anxieux, ainsi qu'une augmentation de l'affect positif et de la satisfaction de vie. (p. 245)</i>	Aucune amélioration des résultats psychosociaux n'est observée chez les garçons à la naissance pour les variables rapportées.
<i>Les augmentations de la congruence de l'apparence étaient associées à une diminution des symptômes de dépression et d'anxiété, et à une augmentation de l'affect positif et de la satisfaction de vie. (p. 245)</i>	Les résultats montrent que congruence d'apparence n'est pas associée à une amélioration des résultats psychosociaux chez les hommes à la naissance.
<i>(...) nos résultats ont montré une amélioration du fonctionnement psychosocial sur deux années de traitement hormonal affirmant le genre, ce qui soutient l'utilisation des GAH comme traitement efficace pour les jeunes</i>	Comme nous le verrons à la section suivante, aucun élément ne permet d'attribuer les évolutions observées aux hormones seules. L'effet principal concerne uniquement la congruence d'apparence.

<i>transgenres et non binaires. (...)</i> <i>Globalement, nos résultats apportent des éléments en faveur de l'efficacité des GAH pour améliorer la congruence de l'apparence et le fonctionnement psychosocial chez les jeunes transgenres et non binaires. . (p. 249)</i>	
---	--

Cette analyse montre que les auteurs formulent des **affirmations qui ne sont pas étayées par leurs propres résultats**, ce qui constitue un **biais de présentation manifeste**, également désigné sous le terme de *spin*. Ce type de biais se caractérise par des stratégies de présentation qui ont pour effet de «*dénaturer l'interprétation des résultats et induire les lecteurs en erreur afin que ceux-ci soient perçus de manière plus favorable* » (Chiu et al., [2017](#)).

Comme le montre le tableau 5, le biais de présentation opère à plusieurs niveaux :

- par la sélection des seuls résultats psychosociaux favorables ;
- par la généralisation excessive d'effets observés uniquement chez les filles assignées filles à la naissance ;
- par l'affirmation non étayée d'un lien entre la congruence d'apparence et l'amélioration psychosociale chez l'ensemble des participants ;
- et par une présentation causale de changements qui ne peuvent être attribués au traitement en l'absence de groupe contrôle ou d'ajustement des facteurs de confusion.

En donnant une **image trompeuse** de l'efficacité du traitement, ces déclarations exposent le lecteur à une compréhension erronée des données, avec des **conséquences cliniques potentielles** qui seront discutées ultérieurement.

Enfin, il est notable que la figure S3 (déjà mentionnée), qui présente l'évolution temporelle des cinq variables psychosociales rapportées selon le sexe, ait été relégué dans l'[annexe](#). Dans une étude prospective longitudinale, ces trajectoires sont pourtant centrales. Son exclusion du corps de l'article nuit à la lisibilité globale des résultats et pose question sur les choix de mise en avant opérés.

4. Limites majeures dans l'attribution causale des effets observés

Au-delà du cas particulier de la congruence d'apparence, l'étude souffre de limitations méthodologiques importantes qui empêchent d'attribuer avec rigueur les évolutions psychosociales observées au traitement hormonal lui-même.

Les auteurs rapportent que certains participants présentant initialement des niveaux modérés ou sévères de dépression ou d'anxiété ont vu leurs scores s'améliorer au point de sortir de la zone clinique : par exemple concernant la dépression, ils rapportent que parmi les 217 participants ayant été évalués sur la dépression au début de l'étude et à 24 mois, « *Trente-cinq participants présentaient des scores de dépression dans la catégorie sévère au départ ; la*

majorité d'entre eux (20 [57,1 %]) ont rapporté un score de dépression relevant des catégories minimale ou modérée après 24 mois ». (p. 243)

Ces cas individuels peuvent sembler encourageants, mais ils appellent à une interprétation prudente pour plusieurs raisons :

- L'étude ne comporte **aucun groupe de comparaison**. Il est donc impossible de distinguer les effets du traitement hormonal de ceux d'autres facteurs externes (maturation psychologique, soutien familial, évolution spontanée de la détresse, etc.).
- **Les traitements concomitants** n'ont pas été intégrés dans les analyses. Par exemple, la variable « médication psychotrope » est bien mentionnée dans les dossiers médicaux (Chen et al., [2021](#), p. 10/44), mais n'est pas incluse dans les modèles statistiques. Or, il est plausible que des participants souffrant de dépression aient reçu des antidépresseurs, dont l'effet attendu est précisément la réduction des symptômes dépressifs. Le même raisonnement s'applique au **soutien psychothérapeutique**, également probable dans ce contexte clinique.
- Une part importante des participants initialement en détresse reste au-dessus du seuil clinique après deux ans. Ce constat, conjugué à des tailles d'effet globalement modestes (Cohen's $d \approx 0,20$), suggère que l'amélioration globale du bien-être est limitée.

En l'absence d'un design contrôlé et d'une prise en compte rigoureuse des facteurs de confusion, les améliorations rapportées ne peuvent être considérées comme la démonstration d'un effet spécifique des hormones. L'étude ne permet pas d'établir de lien de causalité, mais seulement des associations descriptives, dont l'interprétation nécessite une prudence particulière.

5. Biais de confirmation dans l'interprétation des résultats négatifs

Afin d'expliquer l'absence d'amélioration des résultats psychosociaux chez les garçons à la naissance, les auteurs avancent deux hypothèses principales :

a. Les effets corporels différés

Les effets physiques des hormones sexuelles croisées pourraient nécessiter entre deux et cinq ans pour se manifester pleinement, tandis que ceux de la testostérone sur les femmes natales seraient plus rapides. Ils suggèrent donc qu'un suivi plus long pourrait être nécessaire pour observer des effets positifs sur les variables psychosociales.

→ Cette explication présente une incohérence manifeste : comme mentionné précédemment, la congruence de l'apparence (censée refléter directement les changements corporels) s'est améliorée de manière similaire chez les garçons et chez les filles. Cela indique que des modifications physiques perceptibles sont déjà survenues, remettant **ainsi en question la validité** de l'explication avancée.

b. Le stress minoritaire différentiel

Les différences observées pourraient être liées à des niveaux de stigmatisation sociale distincts, les personnes transféminines rencontrant davantage de discrimination que les transmasculines (Delozier et al., [2020](#) ; Poquiz et al., [2021](#)).

→ Cette hypothèse n'est étayée par aucune donnée empirique spécifique dans l'étude : aucun indicateur de stress minoritaire ou de stigmatisation n'a été mesuré ni intégré dans les analyses statistiques. Surtout, elle élude la question centrale : les améliorations observées sont-elles réellement dues aux hormones, ou bien à des facteurs sociaux ou psychologiques non modélisés ?

L'ensemble du raisonnement semble témoigner d'une **réticence à envisager une explication moins favorable à l'hypothèse principale** : que les hormones sexuelles croisées pourraient ne pas améliorer significativement le fonctionnement psychosocial des garçons à la naissance. En avançant des hypothèses alternatives non vérifiées (délais corporels, effets sociaux), les auteurs éludent l'absence de bénéfices démontrés. Cette posture interprétative aboutit à une lecture optimiste du traitement, non étayée par les données, et contribue à une présentation biaisée de l'étude, aux implications cliniques potentiellement majeures (voir section suivante).

6. Risques clinique liés et portée élargie d'une interprétation trompeuse

6.1. Une présentation biaisée aux conséquences concrètes

Comme analysé précédemment, les conclusions des auteurs sur l'efficacité du traitement ne sont pas étayées par leurs propres résultats. En l'absence de mise en garde sur les limites méthodologiques de l'étude, cette présentation optimiste peut induire en erreur les professionnels de santé, les patients et les décideurs publics.

Un praticien pourrait ainsi, à tort, conclure que les hormones sexuelles croisées améliorent significativement le fonctionnement psychosocial des jeunes transgenres.

Or, dans le cadre du consentement éclairé, il est essentiel que les professionnels transmettent aux jeunes et à leurs familles des informations équilibrées, transparentes et fondées sur des données solides. En amplifiant les bénéfices supposés et en minimisant les incertitudes ou effets indésirables, ce type de publication nuit à la qualité de la décision médicale, en entravant un consentement réellement éclairé.

6.2. Une surconfiance injustifiée dans un contexte de faible niveau de preuve

Cette problématique est d'autant plus préoccupante que plusieurs évaluations récentes, notamment le Cass Review ([2024](#)) et la revue systématique de Miroshnychenko et al. ([2025](#)), soulignent que les bénéfices psychologiques du traitement hormonal chez les mineurs reposent sur des preuves de très faible certitude.

Dans ce contexte d'incertitude scientifique, l'étude de Chen et al., loin d'éclairer le débat, contribue à entretenir une confiance excessive dans une intervention dont les effets restent largement indémontrés.

6.3. Une tendance préoccupante dans la littérature scientifique

L'étude de Chen et al. s'inscrit dans une tendance plus large de production scientifique favorable à la transition médicale des mineurs, amorcée par des études fondatrices méthodologiquement fragiles (de Vries et al., [2011](#), [2014](#)). Plusieurs travaux ultérieurs (Costa et al. [2015](#), Carmichael et al. [2021](#), Tordoff et al. [2022](#), Turban et al. [2020](#), [2022](#)) ont reproduit des

faiblesses méthodologiques : absence de groupe contrôle rigoureux, interprétations causales discutables, ou présentation sélective des résultats, tout en affirmant des bénéfices marqués.

Cette orientation répétée, mise en évidence par de nombreux auteurs critiques (Abruzzese et al., [2023](#) Biggs [2020](#), [2022](#)), contribue à normaliser des conclusions encore incertaines.

Conclusion

L'étude de Chen et al. ([2023](#)) présente de graves limitations qui compromettent la validité de ses conclusions. Les auteurs ont révisé leurs hypothèses après observation des résultats, sans en faire état explicitement, ce qui soulève un problème majeur de transparence scientifique. Cette révision a conduit à écarter trois quarts des variables clés spécifiées dans le protocole, bien qu'elles aient été mesurées, et à introduire trois nouvelles variables.

Malgré ces écarts substantiels, les auteurs déclarent garantir la fidélité de l'étude au protocole. L'article se trouve entièrement centré sur une variable secondaire, absente du protocole enregistré, tandis que les résultats négatifs sont minimisés, et que les auteurs formulent des affirmations générales non étayées par leurs propres données : un cas typique de *spin bias*, qui nuit à l'interprétation rigoureuse des résultats.

Les améliorations des résultats psychosociaux rapportées chez les filles natales sont faibles et d'une pertinence clinique limitée, tandis qu'aucune amélioration n'est observée chez les garçons. Par ailleurs, **aucune conclusion causale ne peut être valablement tirée** de ces résultats, en raison de l'absence de groupe contrôle et de la non-prise en compte de facteurs confondants.

En raison de l'ensemble de ces limitations méthodologiques, Cheung et al. ([2024](#)) ont classé cette étude dans la catégorie des risques critiques de biais à l'aide de l'outil ROBINS-I.

Ils l'ont également évaluée avec l'échelle de Newcastle–Ottawa (NOS), qui lui a attribué une qualité dite « modérée » ; cependant, cette grille d'évaluation ne prend pas en compte certains biais méthodologiques majeurs, tels que la reformulation post hoc des hypothèses (HARKing) ou les biais interprétatifs.

Cette publication s'inscrit dans une littérature plus large présentant des caractéristiques similaires. Les revues systématiques successives (Koener et al. [2024](#)) ont souligné la faiblesse du niveau de preuve ; la plus récente, Miroshnychenko et al. ([2025](#)), a classé l'ensemble des résultats psychosociaux, y compris ceux issus de l'étude de Chen et al., à un **niveau de certitude « très faible »**.

Cette fragilité des résultats est à mettre en balance avec les **risques potentiels** liés au traitement hormonal à l'adolescence, et à vie : risques osseux et cardiovasculaires (Chan Swe et al., [2022](#)), impact sur la fertilité future (Stolk et al., [2023](#)), dysfonctions sexuelles (Kronthaler et al., [2024](#) ; da Silva et al., [2024](#)), ainsi que des effets encore inconnus à long terme (Cass Review, [2024](#), § 90).

Le tout prescrit à des jeunes dont l'identité et les priorités sont susceptibles d'évoluer au fil du temps (Cass Review, [2024](#), §16.10 ; Bachmann et al., [2024](#) ; Rawee et al., [2024](#) ; Sapir, [2024](#)).

Tant que les données restent aussi fragiles, il est impératif d'adopter une approche prudente, fondée sur des preuves robustes, avant de généraliser ou de médicaliser davantage les trajectoires de jeunes en pleine construction identitaire.

Comme le rappelle l'ESCAP (*European Society for Child and Adolescent Psychiatry*), les résultats de recherche devraient être publiés **uniquement sur la base de critères de qualité**, et non en fonction de leurs conclusions (ESCAP Statement, [2024](#)).

Cette exigence fondamentale semble ici avoir été négligée, tant dans la conduite de l'étude que de son évaluation par les pairs.

Références