



Disponible en ligne sur

ScienceDirect  
www.sciencedirect.com

Elsevier Masson France

EM|consulte  
www.em-consulte.com



Article original

## Recommandations pour la pratique clinique du CNGOF : évaluation un an après révision de la méthodologie

### CNGOF clinical practice guidelines: Evaluation one year after revision of the methodology

J.-L. Brun<sup>a,b,\*</sup>, L. Sentilhes<sup>a</sup>, A. Torre<sup>c</sup>, C. Huchon<sup>d,e</sup>, C. Garabedian<sup>f,g</sup>, G. Legendre<sup>h,i</sup>,  
J. Sibiude<sup>j,k</sup>, M.-V. Sénat<sup>i,l</sup>, H. Marret<sup>m,n</sup>, T. Schmitz<sup>o,p</sup>

<sup>a</sup> Pôle d'obstétrique-reproduction-gynécologie, centre Aliénor d'Aquitaine, hôpital Pellegrin, CHU Bordeaux, place Amélie Raba Léon, 33076 Bordeaux, France

<sup>b</sup> UMR 5234 (MFP), microbiologie fondamentale et pathogénicité, université de Bordeaux, 33076 Bordeaux, France

<sup>c</sup> Centre de procréation médicalement assistée, centre hospitalier Sud Francilien, 40, avenue Serge Dassault, 91106 Corbeil-Essonnes, France

<sup>d</sup> Service de gynécologie-obstétrique, hôpital Lariboisière, AP-HP, 2, rue Ambroise Paré, 75010 Paris, France

<sup>e</sup> UMR 1123 (ECEVE), épidémiologie clinique, évaluation économique, populations vulnérables, université de Paris, Paris, France

<sup>f</sup> Clinique d'obstétrique, hôpital Jeanne de Flandre, CHU de Lille, avenue Eugène Avinée, 59000 Lille, France

<sup>g</sup> ULR 2694 (METRICS), évaluation des technologies de santé et des pratiques médicales, université de Lille, 59000 Lille, France

<sup>h</sup> Service de gynécologie-obstétrique, CHU d'Angers, 4, rue Larrey, 49933 Angers, France

<sup>i</sup> UMR 1018 (CESP), centre de recherche en épidémiologie et santé des populations, 94076 Villejuif, France

<sup>j</sup> Service de gynécologie-obstétrique, hôpital Louis Mourier, AP-HP, 92700 Colombes, France

<sup>k</sup> UMR 1137 (IAME), centre de recherche infection-antimicrobiens-modélisation-évolution, université de Paris, 75018 Paris, France

<sup>l</sup> Service de gynécologie-obstétrique, hôpital Bicêtre, AP-HP, 78, avenue du Général Leclerc, 94270 Le Kremlin-Bicêtre, France

<sup>m</sup> Service de chirurgie pelvienne gynécologique et oncologique, hôpital Bretonneau, CHRU de Tours, 2, boulevard Tonnellé, 37044 Tours, France

<sup>n</sup> UMR 1253, Imagerie et cerveau, université de Tours, 37032 Tours, France

<sup>o</sup> Service de gynécologie-obstétrique, hôpital Robert-Debré, AP-HP, 48, boulevard Sérurier, 75019 Paris, France

<sup>p</sup> INSERM U1153 (EPOPé), épidémiologie obstétricale périnatale et pédiatrique, université de Paris, 75006 Paris, France

#### INFO ARTICLE

Historique de l'article :

Reçu le 7 novembre 2021

Accepté le 16 novembre 2021

Disponible sur Internet le 18 novembre 2021

Mots clés :

Recommandations pour la pratique clinique

Consortium AGREE II

Question PICO

Système GRADE

#### R É S U M É

**Objectifs.** – Évaluer la révision méthodologique des recommandations pour la pratique clinique (RPC) du collège national des gynécologues obstétriciens français (CNGOF).

**Méthode.** – Trois RPC ont été organisées en 2020 sur les thèmes de la pré-éclampsie sévère, des ménorragies et de la chirurgie prophylactique selon AGREE II (Appraisal of Guidelines for Research & Evaluation). Les questions ont été présentées selon le format PICO (Population, Intervention, Comparison, Outcome) et la cotation de la preuve scientifique selon la méthode GRADE (Grading of Recommendation Assessment, Development and Evaluation).

**Résultats.** – Les trois groupes RPC ont adhéré à cette nouvelle méthodologie. Cependant la présentation des argumentaires, de la formulation des recommandations et de l'élaboration des tableaux GRADE étaient hétérogènes d'un groupe à l'autre. Une homogénéisation de la présentation est proposée, ainsi qu'un guide à l'analyse critique de la littérature pour aider les experts à la cotation de la preuve.

**Conclusion.** – L'adhérence à ces critères de qualité devrait permettre une application plus facile des recommandations au plan national et une meilleure reconnaissance internationale des travaux réalisés par le CNGOF.

© 2021 Elsevier Masson SAS. Tous droits réservés.

\* Auteur correspondant.

Adresse e-mail : [jean-luc.brun@chu-bordeaux.fr](mailto:jean-luc.brun@chu-bordeaux.fr) (J.-L. Brun).

## A B S T R A C T

**Keywords:**

Clinical practice guideline  
 AGREE II consortium  
 PICO question  
 GRADE system

**Objectives.** – To evaluate the revision of methodology of the clinical practice guidelines (CPG) of the French National College of Gynecologists and Obstetricians (CNGOF).

**Method.** – Three CPGs were organized in 2020 on the topics of severe preeclampsia, menorrhagia, and prophylactic surgery according to AGREE II (Appraisal of Guidelines for Research & Evaluation). Questions were presented in PICO (Population, Intervention, Comparison, Outcome) format and the grading of scientific evidence was based on the GRADE (Grading of Recommendation Assessment, Development and Evaluation) method.

**Results.** – All three CPGs groups adhered to this new methodology. However, the presentation of the arguments, the formulation of the recommendations and the development of the GRADE tables were heterogeneous from one group to another. A homogenization of the presentation is proposed, as well as a guide to the critical analysis of the literature to help the experts to rate the evidence.

**Conclusion.** – Adherence to these quality criteria should make it easier to apply the recommendations at the national level and improve international recognition of the work done by the CNGOF.

© 2021 Elsevier Masson SAS. All rights reserved.

## 1. Introduction

La révision de l'organisation et de la méthodologie des recommandations pour la pratique clinique (RPC) par le collège national des gynécologues obstétriciens français (CNGOF) a conduit à abandonner le système proposé par la Haute Autorité de Santé (HAS) pour adopter les critères du consortium AGREE II (Appraisal of Guidelines for Research & Evaluation). Le but est d'améliorer la qualité des recommandations, permettre une meilleure assimilation des messages clés et réduire les écarts existants entre les recommandations et les pratiques cliniques de terrain [1].

Trois RPC ont été organisées en 2020 selon cette nouvelle méthodologie sur les thèmes de la pré-éclampsie sévère, des ménorragies et de la chirurgie prophylactique ; elles ont été présentées aux journées Paris Santé Femmes (PSF) en 2021.

Le conseil scientifique du CNGOF s'est réuni à deux reprises en 2021 pour évaluer le résultat de ces RPC et proposer d'uniformiser la présentation des argumentaires et des tableaux permettant l'élaboration des recommandations.

## 2. Méthodologie AGREE II

L'adoption par le CNGOF des critères de qualité AGREE II va permettre à nos RPC d'être plus facilement identifiées comme répondant aux standards actuels de qualité [2]. La grille d'évaluation est composée de 23 critères de qualité répartis en 6 domaines (Tableau 1). Le domaine 1 mentionne que les questions cliniques faisant l'objet de la RPC doivent être décrites explicitement ; c'est pour cela qu'elles sont désormais présentées selon le format PICO (Population, Intervention, Comparison, Outcome) comme

**Tableau 1**

Grille d'évaluation de la qualité des recommandations pour la pratique clinique (AGREE II).

<b>Domaine 1. Champ et Objectifs</b>	
1.	Le ou les objectifs de la RPC sont décrits explicitement
2.	La ou les questions cliniques couvertes par la RPC sont décrites explicitement
3.	La population (patients, public, etc.) à laquelle la RPC doit s'appliquer est décrite explicitement
<b>Domaine 2. Participation des groupes concernés</b>	
4.	Le groupe ayant élaboré la RPC inclut des représentants de tous les groupes professionnels concernés
5.	Les opinions et les préférences de la population cible (patients, public, etc.) ont été identifiés
6.	Les utilisateurs cibles de la RPC sont clairement définis
<b>Domaine 3. Rigueur d'élaboration de la RPC</b>	
7.	Des méthodes systématiques ont été utilisées pour rechercher les preuves scientifiques
8.	Les critères de sélection des preuves sont clairement décrits
9.	Les forces et les limites des preuves scientifiques sont clairement décrites
10.	Les méthodes utilisées pour formuler les recommandations sont clairement décrites
11.	Les bénéfices, les effets secondaires et les risques en termes de santé ont été pris en considération dans la formulation des recommandations
12.	Il y a un lien explicite entre les recommandations et les preuves scientifiques sur lesquelles elles reposent
13.	La RPC a été revue par des experts externes avant sa publication
14.	Une procédure d'actualisation de la RPC est décrite
<b>Domaine 4. Clarté et présentation</b>	
15.	Les recommandations sont précises et sans ambiguïté
16.	Les différentes options de prise en charge de l'état ou du problème de santé sont clairement présentées
17.	Les recommandations clés sont facilement identifiables
<b>Domaine 5. Applicabilité</b>	
18.	La RPC décrit les éléments facilitant son application et les obstacles
19.	La RPC offre des conseils et/ou des outils pour mettre les recommandations en pratique
20.	Les répercussions potentielles de l'application des recommandations sur les ressources ont été examinées
21.	La RPC propose des critères de suivi et/ou de vérification
<b>Domaine 6. Indépendance éditoriale</b>	
22.	Le point de vue des organismes de financement n'a pas influencé le contenu de la RPC
23.	Les intérêts divergents (les conflits d'intérêts) des membres du groupe ayant élaboré la RPC ont été pris en charge et documentés

recommandé par l'OMS [3]. Les domaines 3 et 4 portant sur la rigueur d'élaboration et sur la clarté de la présentation nous ont conduits à utiliser une grille de cotation de la preuve scientifique selon la méthode GRADE (Grading of Recommendation Assessment, Development and Evaluation) [4].

### 2.1. Questions PICO

La formulation des questions PICO constitue la première étape de l'élaboration d'une RPC. Elle permet de définir une population ciblée et d'analyser l'impact de l'intervention par rapport à un comparateur sur un critère de jugement donné. Ainsi dans une population donnée (P), le résultat d'un traitement (I) est évalué par rapport à un traitement de référence (C), voire l'absence de traitement, sur un critère de jugement donné tel qu'un résultat clinique ou paraclinique (exemple dans le Tableau 2). La formulation permet aussi de comparer la performance d'un test diagnostique (I) par rapport à un test diagnostique de référence (C) ou éventuellement l'absence de test, sur un critère de jugement (O) pouvant correspondre au diagnostic de la maladie (exemple dans le Tableau 3).

La formulation d'une question au format PICO implique de sélectionner les critères de jugement a priori, avant l'analyse de la littérature, puis de la classer selon leur importance. Les critères de jugement choisis doivent être communs aux différentes publications de façon à répondre à la question posée et permettre de réaliser plus aisément la synthèse. Il existe une classification allant de 1 à 9, mais en pratique, on les définit selon trois classes (cruciale, importante, pas importante). Il est conseillé de choisir 2 critères de jugement en privilégiant les classes plus élevées. Si la littérature est de qualité et permet d'évaluer les critères de jugement cruciaux ou élevés, il n'est pas nécessaire d'en étudier d'autres. Chaque critère de jugement est évalué par un ou plusieurs critères d'analyse. Il est également utile de les classer, le premier critère d'analyse étudié étant le critère principal. À titre d'exemple en obstétrique, pour une question portant sur l'efficacité du traitement de la menace d'accouchement prématuré avec comme critère de jugement la morbidité néonatale (critère crucial), le critère d'analyse principal serait la mortalité néonatale et les autres critères seraient l'hospitalisation en réanimation néonatale, le score d'Apgar inférieur à une valeur donnée, etc. En gynécologie pour une question portant sur l'efficacité de l'endométréctomie sur les ménorragies avec comme critère de jugement le score des saignements (critère crucial), le critère d'analyse principal serait l'aménorrhée car les études sont souvent construites pour répondre à cet objectif. Cependant en pratique clinique le but est aussi d'obtenir un contrôle des saignements excessifs (oligoménorrhée, euménorrhée), qui sont les autres critères d'analyse à considérer. La satisfaction des patientes et l'absence de chirurgie itérative sont les autres critères de jugement (critères importants) de l'efficacité de l'endométréctomie. Enfin, les symptômes et la qualité de vie peuvent être évalués sur des critères d'analyse composites établis à partir de questionnaires validés, tel que UFS-QOL pour les myomes ; ce questionnaire intègre des données sur les ménorragies, mais leur extraction peut se révéler difficile pour répondre aux critères de jugement.

Pour chaque question PICO, un court argumentaire est d'abord rédigé selon les données de la littérature, puis, en fin de question, une synthèse et une recommandation (si cela est possible) sont élaborées selon la qualité de la preuve.

### 2.2. Tableaux GRADE

La méthode GRADE permet de coter la qualité de la preuve scientifique selon les données de la littérature en réponse à la question PICO. Cette méthode est largement adoptée par un grand

nombre de sociétés savantes internationales de diverses spécialités depuis une vingtaine d'années [4].

L'ancienne méthodologie HAS corrélait la force de la recommandation au niveau de preuve scientifique [5]. Il y a plus de nuances avec la méthode GRADE qui introduit désormais la notion de qualité de preuve, préférée au niveau de preuve, en tenant compte par exemple des biais potentiels, de la reproductibilité et de l'amplitude de l'effet ou des facteurs de confusion. Les études ne sont plus considérées indépendamment en fonction de leur propre niveau de preuve ; seule la synthèse de la littérature, répondant aux critères de jugement de la question PICO, est rapportée avec une qualité de preuve globale.

Ce travail de synthèse est résumé dans des tableaux reprenant l'ensemble de la littérature, chaque étude occupant une ligne. La valeur initiale de la qualité de la preuve est élevée (cotée +4) lorsqu'il s'agit d'essais randomisés ou basse (cotée +2) lorsqu'il s'agit d'études observationnelles. La méthode GRADE permet de diminuer ou d'augmenter la valeur de la qualité de preuve initiale selon plusieurs critères qui permettent de définir le score final : élevé (+4), modéré (+3), bas (+2), ou très bas (+1). La valeur de la qualité de la preuve permet d'établir le niveau final de confiance que l'on peut avoir vis-à-vis de l'effet d'une stratégie diagnostique ou thérapeutique étudiée par rapport à la référence dans une population donnée pour un critère de jugement défini. Ainsi, pour une qualité de preuve élevée, des études futures auront peu de chances de modifier notre confiance dans l'estimation de l'effet ; à l'inverse pour des qualités de preuves très basses ou basses, l'estimation de l'effet est très incertaine et des études futures auront très probablement un impact important sur notre confiance dans l'estimation de l'effet.

### 2.3. Recommandations

La formulation d'une recommandation est établie à partir de la qualité de la preuve résultant des tableaux GRADE, mais également d'autres éléments tels que l'importance de la pathologie, les valeurs et les préférences des soignants et des patientes, la balance bénéfice risque, l'acceptabilité, la faisabilité, le coût et l'équité sociale.

Les recommandations ne comprennent qu'une dimension positive du type « il est recommandé de... » ou négative « il est recommandé de ne pas... » avec une force faible ou forte. Le sens de la recommandation est donné dans son intitulé associé à sa force et sa qualité de la preuve (Tableaux 2 et 3). Dans un certain nombre de situations où la littérature scientifique est insuffisante, les questions ne trouvent pas de réponse. La littérature peut aussi donner des résultats divergents. La synthèse de l'argumentaire permet de préciser l'une ou l'autre de ces situations avec la conclusion suivante : pas de recommandation.

## 3. Évaluation des trois RPC de 2020

Les trois RPC de 2020 ne sont pas déroulées de la même manière, car la RPC sur la pré-éclampsie sévère a été réalisée en collaboration avec la Société Française d'Anesthésie Réanimation, contrairement aux deux autres où les groupes de travail étaient autonomes et dont la constitution a été proposée et validée par le conseil d'administration et le conseil scientifique du CNGOF.

Les objectifs ont été atteints dans ces trois RPC. Des recommandations positives ou négatives, ou l'absence de recommandation ont pu être établies à partir d'argumentaires basés sur des tableaux GRADE pour des questions PICO bien formulées. Cependant, quelques accords professionnels ont été formulés, ce qu'il est souhaitable d'éviter car relevant plus d'un consensus formalisé d'experts que de RPC [1]. Par ailleurs, la présentation des

**Tableau 4**

Question PICO de type diagnostique ou thérapeutique : modèle de présentation retenu (format A4). **Chez les femmes d'une population P, l'examen/traitement étudié I est-il plus performant/efficace que l'examen/traitement de référence C pour évaluer/traiter le critère O ?**

**Introduction sur l'importance de la PICO (5 à 10 lignes)**

La pathologie et les stratégies comparées

**Argumentaire (1 à 1,5 pages)**

Rapporter les résultats des études mentionnées dans les tableaux GRADE et leur qualité de preuve.

Intégrer l'acceptabilité et la faisabilité des stratégies étudiées, la balance bénéfique risque et le cout, les préférences des soignants ou des patientes selon la question posée.

**Synthèse** : elle doit répondre à la question mais ce n'est pas encore une recommandation.

Exemple 1 : chez les femmes d'une population P, l'examen I est plus performant que l'examen C pour évaluer le critère O.

Exemple 2 : chez les femmes d'une population P, le traitement I est aussi efficace et aussi bien toléré que le traitement C pour traiter O.

Exemple 3 : en l'absence/insuffisance des données de la littérature, ou compte tenu des données contradictoires de la littérature, il n'y a pas d'argument pour proposer préférentiellement un examen/traitement I ou C pour permettre le diagnostic/traitement de O.

**R1 – Chez une femme d'une population P, il est recommandé de réaliser l'examen I plutôt que l'examen C pour diagnostiquer O.**

**Recommandation faible/forte, Qualité de preuve élevée/modérée/basse/très basse**

**R2 – Chez une femme d'une population P, il est recommandé indifféremment le traitement I ou C pour traiter O.**

**Recommandation faible/forte, Qualité de preuve élevée/modérée/basse/très basse**

**R3 – Pas de recommandation**

**Références** listées par ordre de citation selon format Vancouver police arial 8

tableaux GRADE et des argumentaires aux questions PICO était différente entre les groupes de travail.

Après avoir comparé les modes de présentation des trois RPC rédigées en 2020, les auteurs du texte se sont accordés pour adopter un modèle permettant la meilleure compréhension du texte et une lecture optimale des tableaux.

**3.1. Argumentaire et recommandations**

Le modèle de présentation finalement retenu à la suite des 2 visio-conférences est rapporté dans le **Tableau 4**. L'argumentaire intègre les différents critères associés aux études cliniques et conclut par une synthèse : une phrase positive (l'effet du médicament A est supérieur à celui du médicament B pour traiter une pathologie C), une phrase négative (la performance de

l'examen A est inférieure à celle de l'examen B pour le diagnostic d'une pathologie C) ou une phrase ne permettant pas de conclure (il n'y a pas d'argument pour proposer un traitement par... ; les données de la littérature ne permettent pas de sélectionner un examen diagnostique pour... ; les données de la littérature sont trop limitées en qualité ou en nombre pour pouvoir proposer... ; l'hétérogénéité des données de la littérature ne permet pas de valider une stratégie...). La recommandation est définie en précisant la population ciblée, avec un sens négatif ou positif en proscrivant la formulation du type « il n'est pas recommandé de », autrefois usitée qui pouvait donner lieu à confusion [6]. La recommandation (faible ou forte) est formulée avant la qualité de preuve (très basse, basse, modérée ou élevée). Dans une recommandation positive, deux modalités sont possibles : il est recommandé le traitement A plutôt que le traitement B si la nouvelle intervention est supérieure au gold standard (traitement B dans cet exemple) ; il est recommandé indifféremment le traitement A ou le traitement B pour deux stratégies également répandues qui n'ont pas fait la preuve d'une supériorité l'une par rapport à l'autre. En l'absence de conclusion possible basée sur les preuves, la mention « pas de recommandation » sera préférée à un « accord professionnel ».

Les références sont mentionnées dans chaque PICO selon le format Vancouver par ordre de citation, en police 8.

**3.2. GRADE**

L'analyse critique de la présentation des données des différents tableaux a permis de sélectionner deux modèles mentionnés dans les **Tableaux 5 et 6 (Tableaux 5 (études comparatives) et 6 (études non comparatives))**. Trois colonnes majeures se composent de plusieurs sous-colonnes de la façon suivante : description de l'étude (auteur et année de publication, plan d'expérience et population, critères d'analyse) ; effet (intervention/contrôle ou patientes, résultats) ; preuve (niveau initial, décote, surcote, qualité de preuve finale).

Concernant l'effet et les résultats, les données qualitatives sont présentées en valeur numérique (nominator/dénominateur) avec pourcentage (%) et les données quantitatives selon la valeur médiane encadrée des valeurs minimales et maximales. La 1<sup>re</sup> valeur concerne l'intervention (I) et la 2<sup>e</sup> valeur est le contrôle (C). Il est préférable de respecter une présentation homogène de ces valeurs pour éviter les confusions. La 1<sup>re</sup> ligne des résultats du **Tableau 7** mentionne un risque relatif ou un *odds-ratio* (avec son intervalle de confiance à 95 %) ou une différence de moyenne avec un test statistique permettant d'établir la valeur du P sur la 2<sup>e</sup> ligne.

**Tableau 5**

Modèle à utiliser dans les études comparatives où les résultats de l'intervention sont décrits en comparaison à un contrôle.

Critère de jugement: (crucial, important, pas important)										
Description de l'étude				Effet			Preuve			
Auteur (Année)	Type d'étude	Plan d'expérience Population	Critères d'analyse	Intervention	Contrôle	Résultats (IC95%)	NI	Décote	Surcote	QP
	Méta analyse d'ECR Essai randomisé Cohortes	Plan Critères d'inclusion	Critère 1*	X/XX (%) ou Médiane (min-max)	Y/YY (%) ou Médiane (min-max)	RR=X,X (Y,Y-Z,Z) ou différence de moyenne Valeur du P En faveur de/défaveur de/pas de différence significative	4	Biais <sup>1</sup> Précision <sup>1</sup> Consistance <sup>1</sup> Applicabilité <sup>1</sup>	Amplitude de l'effet <sup>1</sup> Effet dose <sup>1</sup> Facteur de confusion <sup>1</sup>	1 2 3 4
				Critère 2	X/XX (%) ou Médiane (min-max)	Y/YY (%) ou Médiane (min-max)	RR=X,X (Y,Y-Z,Z) ou Différence de moyenne Valeur du P En faveur de/défaveur de/pas de différence significative			

NI : niveau initial; QP : qualité de preuve; \* Critère principal.

<sup>1</sup>Notes de bas de tableaux détaillant les raisons de décote ou de surcote.

**Tableau 6**

Modèle à utiliser dans les études de cohortes non comparatives où les résultats de l'intervention sont seulement décrits sans comparaison à un contrôle.

Critère de jugement: (crucial, important, pas important)									
Description de l'étude				Effet		Preuve			
Auteur (Année)	Type d'étude	Plan d'expérience Population	Critères d'analyse	Patientes	Résultats (IC95%)	NI	Décote	Surcote	QP
	Méta analyse de cohortes	Plan	Critère 1 *	N	X/XX (%) ou médiane (min-max)	2	Biais <sup>1</sup>	Amplitude de l'effet <sup>1</sup>	1
	Cohortes	Critères d'inclusion			Si comparaison avant/après traitement : différence de fréquence / moyenne Valeur du P En faveur de/défaveur de/pas de différence significative		Précision <sup>1</sup> Consistance <sup>1</sup> Applicabilité <sup>1</sup>	Effet dose <sup>1</sup> Facteur de confusion <sup>1</sup>	2 3 4
			Critère 2		X/XX (%) ou médiane (min-max) Si comparaison avant/après traitement : différence de fréquence / moyenne Valeur du P En faveur de/défaveur de/pas de différence significative	2			

NI : niveau initial; QP : qualité de preuve; \* Critère principal.

<sup>1</sup>Notes de bas de tableaux détaillant les raisons de décote ou de surcote.

**Tableau 7**

Aide méthodologique pour définir les critères de pondération de la qualité de preuve.

Critères de pondération	Causes de variation de confiance en la preuve	Barème
DECOTE		
Biais	Essai randomisé ± méta-analyse Biais de sélection : processus de randomisation ou règles d'attribution non détaillées ou inadéquates; biais de recrutement dans les essais cliniques randomisés par grappe (cluster) Biais d'exécution : aveugle inadéquat ou insuffisant ; persistance de l'effet précédent dans les plans d'étude croisés (essais de type cross-over) Biais d'exclusion : taux important de perdus de vue ; pas d'analyse en intention de traiter Biais de détection : mesures différentes des critères de jugement selon les groupes ; durée de suivi non appropriée ; utilisation d'un critère de jugement non validé ou subjectif Arrêt prématuré de l'étude Biais de publication (multiples études de petite taille aux résultats positifs, possiblement financés par l'industrie) Étude observationnelle ± méta-analyse Biais de sélection pour la répartition des patientes : différences entre les groupes ; pas d'ajustement pour les facteurs confondants Biais de détection : mesures différentes des critères de jugement selon les groupes ; durée de suivi non appropriée Biais d'exclusion : taux important de perdus de vue Biais de publication (multiples études de petite taille aux résultats positifs, possiblement financés par l'industrie)	0 à -2
Précision	Étude non dimensionnée pour répondre à l'objectif principal Petit nombre d'événements Larges intervalles de confiance (différence d'attitude thérapeutique si on considère l'une ou l'autre borne de l'intervalle de confiance) Incertitude sur l'amplitude de l'effet	0 à -1
Consistance	Contradiction inexplicée entre études ou au sein de la même étude Hétérogénéité des résultats dans les méta-analyses	0 (I <sup>2</sup> < 25 %) à -1 (I <sup>2</sup> > 50 %)
Applicabilité	Extrapolation des résultats analysés à la question d'intérêt (objet de la PICO) : hétérogénéité des populations étudiées (âge, sexe etc.), des interventions (dose, protocole etc.), des critères de jugement (méthode de mesure, temps, importance, évaluation intermédiaire) Estimation de l'effet direct ou non : A > B et B > C pour estimer A > C Faible degré de compatibilité entre les variables d'intérêt	0 à -1
SURCOTE		
Amplitude de l'effet	Association forte (RR > 2 ou < 0,5) à très forte (RR > 5 ou < 0,2), à pondérer avec l'intervalle de confiance	0 à +2
Effet dose	Mise en évidence d'un gradient de l'effet en fonction de l'intensité de l'intervention	0 à +1
Prise en compte des facteurs de confusions	Correction des facteurs de confusion minimisant l'effet observé par rapport à l'effet réel même si la prise en compte de ces facteurs rend significatif une association qui initialement ne l'était pas	0 à +1

La 3<sup>e</sup> ligne mentionne le sens du résultat par rapport à l'intervention I (en faveur de I, en défaveur de I ou pas de différence), comme cela est souvent mentionné dans les méta-analyses pour aider à la compréhension.

Le détail des 7 critères de pondération peut être regroupé en 2 colonnes de décote ou de surcote, ce qui améliore la lecture. Les raisons qui ont conduit à décoté la qualité de la preuve (biais, précision, consistance, applicabilité) ou à le surcoté (amplitude de

l'effet, effet dose, facteurs de confusion) doivent être détaillées en notes de bas de tableaux (Tableau 7). Cela permet à chaque expert de préciser aux coordonnateurs de la RPC les raisons qui l'ont conduit à diminuer ou augmenter la note initiale, mais également de faire partager aux lecteurs la cotation de la qualité de preuve à titre didactique. Cet exercice est probablement le temps le plus difficile de l'analyse critique de la littérature. En effet, les risques de biais peuvent être multifactoriels (biais de sélection des patientes,

d'interprétation des données, absence de témoin, biais méthodologique tel qu'un effet indirect lorsque le critère de jugement n'est pas l'objectif principal de l'étude ou lorsque la population étudiée n'est pas la population cible de la recommandation, biais de publication) ; l'imprécision des données peut être établie sur un large intervalle de confiance malgré un risque relatif significatif ; l'inconsistance peut être secondaire aux données hétérogènes inexpliquées de la littérature concernant un critère de jugement donné ; un défaut d'applicabilité ou de validité externe peut être attribué à une méthode d'intervention inadéquate ou non reproductible dans une population donnée. Parmi les facteurs de surcote, l'amplitude de l'effet peut être forte avec un risque relatif  $>2$  ou  $<0,5$  (+1) ou très forte avec un risque relatif  $>5$  ou  $<0,2$  (+2). La mise en évidence d'un effet-dose ou la prise en compte des facteurs de confusion sont également valorisées. Chaque note positive ou négative doit être de 1 voire 2 points au maximum pour éviter des résultats aberrants sur la cotation finale de la qualité de preuve qui doit rester comprise entre +1 (très basse) et +4 (élevée).

La présentation des tableaux selon le format paysage en utilisant une police Arial 9 permet d'introduire un grand nombre d'informations tout en restant lisible. Les deux premières colonnes portant sur la description de l'étude et l'effet sont de dimensions supérieures à la 3<sup>ème</sup> colonne qui permet le calcul de la qualité de preuve.

#### 4. Conclusion

L'ensemble des questions PICO et de leurs argumentaires permettant d'établir les recommandations font l'objet d'une publication unique en français dans la revue Gynécologie obstétrique fertilité sénologie et d'une publication en anglais dans une seule autre revue. Les tableaux GRADE sont indexés en fichiers supplémentaires sur la version numérique. Un exemplaire des tableaux en format Word est à la disposition des experts amenés à participer aux futures RPC en cliquant sur le lien permettant

d'accéder aux fichiers supplémentaires. L'adhérence à ces critères de qualité devrait permettre une application plus facile des recommandations au plan national et une meilleure reconnaissance internationale des travaux réalisés par le CNGOF.

#### Déclaration de liens d'intérêts

Les auteurs déclarent ne pas avoir de liens d'intérêts.

#### Supplément en ligne. Matériel complémentaire

Le matériel complémentaire (Tableaux 2, 3, 5 et 6) accompagnant la version en ligne de cet article est disponible sur <http://www.sciencedirect.com> et <https://doi.org/10.1016/j.gofs.2021.11.008>.

#### Références

- [1] Schmitz T, Senat MV, Sentilhes L, Azria É, Deneux-Tharoux C, Huchon C, et al. *Recommandations pour la pratique clinique du CNGOF: révision méthodologique du processus d'élaboration*. Gynecol Obstet Fertil Senol 2020;48(1):3–11.
- [2] Grille d'évaluation de la qualité des recommandations pour la pratique clinique (grille AGREE II). The AGREE Next Steps Research Consortium; 2009, [https://www.agreetrust.org/wp-content/uploads/2013/06/AGREE\\_II\\_French.pdf](https://www.agreetrust.org/wp-content/uploads/2013/06/AGREE_II_French.pdf).
- [3] Thomas J, Kneale D, McKenzie E, Brennan JE, Bhaumik SES. Chapter 2: Determining the scope of the review and the questions it will address. In: Higgins J, Thomas J, editors. *Cochrane Handbook for Systematic Reviews of Interventions*. Version 6. 2., 2021. <http://www.handbook.cochrane.org>.
- [4] Schönemann H, Brožek J, Guyatt G, Oxman A. *Handbook for grading the quality of evidence and the strength of recommendations using the GRADE approach*. Updated October 2013. <https://gdt.grade.pro.org/app/handbook/handbook.html>.
- [5] Etat des lieux. In: *Niveau de preuve et gradation des recommandations de bonne pratique*. Haute Autorité de Santé; 2013. [https://www.has-sante.fr/jcms/c\\_1600564/fr/niveau-de-preuve-et-gradation-des-recommandations-de-bonne-pratique-etat-des-lieux](https://www.has-sante.fr/jcms/c_1600564/fr/niveau-de-preuve-et-gradation-des-recommandations-de-bonne-pratique-etat-des-lieux).
- [6] Sentilhes L, Schmitz T, Rozenberg P, Verspyck E, Senat MV. *Présentation du siège. Recommandations pour la pratique clinique du CNGOF – Méthode et organisation*. Gynecol Obstet Fertil Senol 2020;48:61–2.