

Les niveaux de preuve

UE E.1 : Recherche, méthodes, analyse critique et données probantes

Classer la force d'une preuve, et comprendre ce que le classement ne dit pas



Elise VEROT

Professeure des Universités en Sciences Infirmières

Directrice du Pôle paramédical et du Département Sciences Infirmières et Techniques en Santé, Faculté de Santé, Université Jean Monnet, Saint-Étienne

Responsable de la recherche paramédicale, CHU de Saint-Étienne
UR 4129 P2S, UCBL1-UJM · CIC 1408 INSERM, CHU de Saint-Étienne
Institut universitaire PRESAGE, UJM

Déroulé du cours E1-N2-03

1. D'où vient l'idée de niveau de preuve
2. Les échelles que vous rencontrerez
3. GRADE, ou le passage du plan à la certitude
4. Ce que la pyramide ne couvre pas



Ce que vous saurez faire à l'issue de ce cours

OUVERTURE

- **Expliquer** d'où vient l'idée de hiérarchiser les preuves, à quelle question elle répond, et énoncer les trois composantes de la pratique fondée sur des **données probantes** ;
- **Lire** une gradation de la Haute Autorité de santé et rattacher un **grade** A, B ou C à son **niveau de preuve** ;
- **Situer** les échelles **OCEBM** et JBI, et dire ce qui les distingue d'une échelle unique ;
- **Expliquer** ce que le système **GRADE** déplace : ce que l'on note n'est plus une étude, mais un corpus, pour un résultat donné ;
- **Nommer** les cinq domaines qui font baisser la certitude et les trois qui la font monter ;
- **Lire** un degré de certitude dans un tableau de **synthèse** de revue ;
- **Décrire** les deux modifications apportées par la pyramide révisée de 2016 ;
- **Dire** pourquoi une étude quasi expérimentale ne se juge pas avec l'outil de l'essai randomisé ;
- **Dire** pourquoi une étude qualitative ne se classe pas sur l'échelle de l'efficacité, et nommer l'échelle qui lui correspond ;
- **Reconnaître** les situations où aucune échelle de niveau ne s'applique (**méthodes mixtes**, recherche en **implémentation**), et argumenter le choix d'un outil d'appréciation à partir de la question posée.

Objectifs déclinés à partir de la fiche UE E.1 (arrêté du 20 février 2026, annexe III)

Lexique des sigles employés dans ce cours

OUVERTURE

- **Systèmes de gradation et échelles de niveau de preuve**
- **GRADE** (Grading of Recommendations Assessment, Development and Evaluation) : système international de gradation de la certitude d'un corpus et de la force des recommandations. **GRADE-CERQual** (Confidence in the Evidence from Reviews of Qualitative research) : son équivalent pour les synthèses qualitatives.
- **HAS** (Haute Autorité de santé) et **ANAES** (Agence nationale d'accréditation et d'évaluation en santé, 1997-2004, à laquelle la HAS a succédé) : gradation française en niveaux 1 à 4 et grades A, B, C. **OCEBM** (Oxford Centre for Evidence-Based Medicine) : échelle d'Oxford 2011. **JB**I (Joanna Briggs Institute) : cinq hiérarchies, une par type de question.
- **Outils d'appréciation d'une étude : un jugement par domaine, sans note globale**
- **RoB 2** (Risk of Bias 2) : essais randomisés. **ROBINS-I** (Risk Of Bias In Non-randomized Studies of Interventions) : interventions non randomisées. **MMAT** (Mixed Methods Appraisal Tool) : études quantitatives, qualitatives et mixtes.
- **Lignes directrices de rédaction : ce qu'un compte rendu doit contenir**
- **COREQ** (Consolidated Criteria for Reporting Qualitative Research) et **SRQR** (Standards for Reporting Qualitative Research) : études qualitatives. **GRAMMS** (Good Reporting of A Mixed Methods Study) : méthodes mixtes.
- **IC 95 %** : intervalle de confiance à 95 %. **RR** : risque relatif. **DMS** : différence moyenne standardisée. **Summary of findings** : tableau de synthèse d'une revue systématique.

Sources : Guyatt et al., 2008 ; Lewin et al., 2015 ; Haute Autorité de santé, 2013 ; Joanna Briggs Institute, 2014 ; Hong et al., 2018

1

D'où vient l'idée de niveau de preuve

Un besoin de tri né de la pratique, avant d'être une échelle

La pratique fondée sur des données probantes

D'OÙ VIENT L'IDÉE

Ce qu'elle est

L'usage consciencieux, explicite et judicieux des meilleures données disponibles pour décider du soin d'un patient donné.

Elle réunit trois choses : les meilleures données externes issues de la recherche, l'expertise clinique du soignant, et les préférences du patient.

Ce qu'elle n'est pas

Ce n'est pas de la cuisine de recette, ni une médecine dictée par des budgets.

Les données externes informent la décision, elles ne la remplacent jamais. Une donnée valide peut être inapplicable ou inacceptable pour un patient précis.

Pourquoi cela oblige à trier

Si l'on doit s'appuyer sur les meilleures données disponibles, encore faut-il pouvoir dire lesquelles sont les meilleures.

Les hiérarchies de preuve sont nées de cette obligation pratique, non d'un goût pour le classement.

Retenez la formule : « les meilleures données disponibles ».

Disponibles, donc parfois faibles. Le niveau de preuve dit où l'on en est, il ne dispense pas de décider.

Sources : Sackett et al., 1996

Le problème que la hiérarchie cherche à résoudre

D'OÙ VIENT L'IDÉE

Deux études qui portent sur la même question peuvent donner deux réponses opposées. Il faut alors un principe pour décider la quelle croire davantage.

L'argument en faveur du classement

Les plans d'étude n'ont pas la même **vulnérabilité aux biais**.

Le tirage au sort équilibre les facteurs **inconnus** ; aucun ajustement statistique ne le fait.

Une échelle donne un **langage commun** entre professionnels et entre disciplines.

Elle rend la recommandation **traçable** : on peut dire sur quoi elle repose.

Ce que le classement ne règle pas

Un essai randomisé **mal conduit** ne vaut pas mieux qu'une bonne cohorte.

Sur plusieurs questions cliniques, essais et études d'observation ont donné des **résultats concordants**.

Le plan seul ne dit rien de la **précision** ni de la **cohérence** des résultats.

Il ne dit rien non plus de la **pertinence** de l'étude pour votre patient.

C'est ce constat qui a conduit, dans les années 2000, à remplacer les échelles fondées sur le seul plan d'étude par des systèmes qui apprécient aussi la conduite de l'étude et la cohérence des résultats.

Sources : Concato et al., 2000 ; Guyatt et al., 2008

La pyramide classique, et ce qu'elle affirme

D'OÙ VIENT L'IDÉE

1

Au sommet : les synthèses

Revue systématiques et méta-analyses d'essais randomisés, présentées comme le degré le plus élevé.

2

Puis les essais randomisés

Le plan qui protège le mieux d'une différence initiale entre les groupes comparés.

3

Puis les études d'observation

Cohortes, puis études cas témoins, puis études transversales et séries de cas.

4

À la base : l'avis d'expert

L'opinion et le raisonnement physiopathologique, tenus pour le degré le plus faible.

Cette forme a une vertu : elle se retient. Elle a un défaut : elle laisse croire que le plan d'étude suffit à décider de la confiance, et qu'une seule échelle vaut pour toutes les questions.

Sources : Murad et al., 2016

La hiérarchie des preuves est un outil de décision, pas un palmarès des chercheurs.

Elle répond à une question précise : à quel point puis-je croire que cet effet est réel ?
Posée à propos d'une intervention, cette question a un sens et une réponse.
Posée à propos d'un vécu de patient, elle n'en a pas.

Sources : Sackett et al., 1996 ; Petticrew & Roberts, 2003

2

Les échelles que vous rencontrerez

France, Oxford, Joanna Briggs : trois manières de classer



La gradation de la Haute Autorité de santé

LES ÉCHELLES EN USAGE

Niveau de preuve scientifique	Types d'études	Grade
Niveau 1	Essais comparatifs randomisés de forte puissance ; méta-analyse d'essais comparatifs randomisés ; analyse de décision fondée sur des études bien menées	A · preuve scientifique établie
Niveau 2	Essais comparatifs randomisés de faible puissance ; études comparatives non randomisées bien menées ; études de cohortes	B · présomption scientifique
Niveau 3	Études cas témoins	C · faible niveau de preuve scientifique
Niveau 4	Études comparatives comportant des biais importants ; études rétrospectives ; séries de cas ; études épidémiologiques descriptives (transversale, longitudinale)	C · faible niveau de preuve scientifique

Rappel de l'ANAES (2000), repris par la HAS (2013) : « Cette gradation des recommandations fondée sur le niveau de preuve scientifique de la littérature venant à l'appui de ces recommandations ne présume pas obligatoirement du degré de force de ces recommandations. En effet, il peut exister des recommandations de grade C ou fondées sur un accord d'experts néanmoins fortes malgré l'absence d'un appui scientifique. »

Sources : Haute Autorité de santé, 2013 ; ANAES, 2000

L'échelle d'Oxford, et l'échelle du Joanna Briggs Institute

LES ÉCHELLES EN USAGE

Ces deux systèmes ont en commun d'avoir renoncé à une échelle unique : ils en proposent une par type de question.

OCEBM (Oxford Centre for Evidence-Based Medicine), 2011

Un tableau à **cinq niveaux**, du plus fort au plus faible.

Il se lit par type de question : fréquence, diagnostic, pronostic, bénéfice du traitement, effets indésirables fréquents ou rares, dépistage.

Il a été conçu comme un **point de départ** pour trouver vite la meilleure donnée, non comme une note de qualité.

JBI (Joanna Briggs Institute), Adélaïde, 2014

Cinq hiérarchies distinctes, une par type de question.

Efficacité, diagnostic, pronostic, évaluation économique, et **signification**.

La hiérarchie « signification » classe les travaux **qualitatifs** : synthèse qualitative, puis étude qualitative unique, puis avis d'expert.

C'est le système le plus explicite sur le fait qu'une échelle unique ne peut pas convenir.

Le JBI est particulièrement présent en sciences infirmières. Sa hiérarchie « signification » reconnaît qu'une étude qualitative n'a pas à être classée sur l'échelle de l'efficacité.

Sources : OCEBM Levels of Evidence Working Group, 2011 ; Joanna Briggs Institute, 2014

Les cinq hiérarchies du JBI, et ce qu'elles montrent

LES ÉCHELLES EN USAGE

Type de question	Ce qui occupe le niveau 1
Efficacité	Revue systématique d'essais randomisés
Diagnostic	Revue systématique d'études de précision diagnostique
Pronostic	Revue systématique de cohortes constituées au début de la maladie (inception cohorts)
Évaluation économique	Modèle de décision alimenté par une revue systématique
Signification	Revue systématique qualitative ou de méthodes mixtes

Lisez la dernière ligne : pour une question de signification, c'est une synthèse qualitative qui occupe le premier rang, et non un essai randomisé. La question commande l'échelle.

Sources : Joanna Briggs Institute, 2014

Il n'existe pas une échelle des preuves, mais des échelles, dont chacune répond à un type de question.

Avant d'appliquer une échelle à un article, demandez-vous quelle question cet article pose. Appliquer l'échelle de l'efficacité à une étude qui n'en pose pas la question produit un verdict qui n'a aucun sens.

Sources : Joanna Briggs Institute, 2014 ; OCEBM Levels of Evidence Working Group, 2011 ; Petticrew & Roberts, 2003

3

GRADE, ou le passage du plan à la certitude

Ce que l'on note n'est plus une étude, mais un corpus



Ce que GRADE déplace

GRADE

GRADE (Grading of Recommendations Assessment, Development and Evaluation) est aujourd'hui le système le plus employé au niveau international, et celui sur lequel s'appuient les revues Cochrane. Il ne remplace pas les échelles précédentes : il change ce que l'on note.

Ce que notaient les échelles classiques

Une **étude**, prise isolément.

Sur son **plan** presque exclusivement.

Une fois pour toutes, quel que soit le résultat considéré.

Le verdict était donc **attaché à l'article**.

Ce que note GRADE

Un **corpus d'études**, pour un **résultat donné**.

Sur son plan, mais aussi sur sa **conduite**, sa **cohérence**, sa **précision** et sa **pertinence**.

Une même revue peut conclure à une certitude élevée sur la mortalité et faible sur la qualité de vie.

Le verdict est donc **attaché à une question**, non à un article.

C'est le point le plus difficile et le plus utile : la certitude ne qualifie pas un article, elle qualifie ce que l'on sait d'un résultat précis à un moment donné.

Sources : Guyatt et al., 2008 ; Schünemann et al., 2024

Rappel : qu'est-ce qu'un biais ?

GRADE

- **Ce que c'est**
- Un biais est une erreur systématique qui écarte le résultat d'une étude de la valeur vraie, dans un sens donné. Il ne dépend pas du hasard : augmenter l'effectif ne le corrige pas, contrairement à l'imprécision.
- **Où il se repère dans un article**
- **Dans la section méthode** : comment les groupes ont été constitués (sélection) ; si les participants et les évaluateurs savaient qui recevait quoi (insu) ; combien de personnes ont été perdues de vue (attrition) ; si tous les résultats prévus ont été rapportés (rapport sélectif).
- **Comment il s'apprécie**
- **Par domaine, avec un outil adapté au plan** : RoB 2 (Risk of Bias 2) pour les essais randomisés, ROBINS-I (Risk Of Bias In Non-randomized Studies of Interventions) pour les interventions non randomisées. Le repérage des biais dans un article est développé dans le cours E1-N2-04 Lire et critiquer un article scientifique ; GRADE en fait le premier de ses cinq domaines.

Sources : Sterne et al., 2019 ; Sterne et al., 2016 ; Schünemann et al., 2024

Les cinq raisons de baisser la certitude

GRADE

Domaine	Ce que l'on constate
Risque de biais	Les études ont des limites de conduite : dissimulation d'attribution insuffisante, absence d'insu, perdus de vue nombreux.
Incohérence	Les études donnent des estimations très différentes, sans explication solide de cette hétérogénéité.
Caractère indirect	La population, l'intervention, le comparateur ou le résultat mesuré s'écartent de ceux de la question posée.
Imprécision	Peu de participants ou peu d'événements, donc un intervalle de confiance large.
Biais de publication	Des études ou des résultats semblent n'avoir pas été publiés en raison de ce qu'ils montraient.

Chaque domaine peut faire descendre d'un ou de deux crans. Un ensemble d'essais randomisés peut ainsi terminer en certitude faible.

Sources : Schünemann et al., 2024 ; Guyatt et al., 2008

Les trois raisons de monter la certitude

GRADE

1

Un effet de grande ampleur

Une estimation d'effet importante, en l'absence de facteurs de confusion plausibles, rend peu vraisemblable que le résultat soit un artefact.

2

Une relation dose-effet

L'effet croît avec le niveau d'exposition. Cette régularité est difficile à produire par la seule confusion.

3

Une confusion qui joue contre l'effet

Tous les biais plausibles auraient dû réduire l'effet observé, et pourtant l'effet est là. Il est donc probablement sous-estimé.

Ces trois raisons ne s'appliquent en pratique qu'aux études d'observation. Elles disent qu'une cohorte peut, dans certaines conditions, soutenir une certitude élevée.

Sources : Schünemann et al., 2024

Les quatre degrés de certitude

GRADE

Degré	Ce qu'il signifie pour le lecteur
Élevée	Nous sommes très confiants que l'effet réel est proche de l'estimation de l'effet.
Modérée	Nous sommes modérément confiants : l'effet réel est probablement proche de l'estimation, mais il pourrait en être nettement différent.
Faible	Notre confiance dans l'estimation est limitée : l'effet réel pourrait être nettement différent de l'estimation.
Très faible	Nous avons très peu confiance dans l'estimation : l'effet réel est probablement nettement différent de l'estimation.

Ces quatre mots sont ceux que vous lirez dans les tableaux de synthèse des revues Cochrane et dans les recommandations internationales. Ils qualifient la confiance, non l'ampleur de l'effet. La formulation antérieure (« de nouvelles recherches modifieront... », Guyatt et al., 2008) se rencontre encore dans des textes d'avant 2011.

Sources : Schünemann et al., 2013 ; Guyatt et al., 2008

Lire un tableau de synthèse (Summary of findings)

GRADE

- Exemple simplifié, construit pour ce cours sur la structure des tableaux Cochrane : une ligne par résultat, la certitude en avant-dernière colonne, et les notes qui disent quel domaine a fait baisser la certitude. Lisez la certitude ligne par ligne : elle qualifie un résultat, non la revue.

Résultat	Effet relatif (IC 95 %)	Participants (études)	Certitude	Notes
Mortalité à 30 jours	RR 0,85 (0,75 à 0,96)	4 200 (8 essais)	Élevée	Aucune baisse : essais bien conduits, résultats concordants, intervalle étroit.
Qualité de vie à 6 mois	DMS 0,20 (−0,05 à 0,45)	610 (3 essais)	Faible	Baissée de deux crans : risque de biais (absence d’insu) ; imprécision (intervalle incluant l’absence d’effet).

Sources : Schünemann et al., 2024 ; Guyatt et al., 2008 (exemple construit pour ce cours)

La pyramide révisée

GRADE

Partant de ce que GRADE a établi, une proposition publiée en 2016 modifie la figure elle-même sur deux points.

Des lignes ondulées, non droites

Les traits qui séparent les étages deviennent **ondulés**.

Motif : le plan d'étude ne suffit pas à apprécier la qualité.

Limites de conduite, imprécision, incohérence et caractère indirect peuvent faire **descendre** un étage.

D'autres facteurs peuvent en faire **monter** un autre.

La revue systématique n'est plus au sommet

Elle est retirée du sommet et devient la **lentille** à travers laquelle on regarde les autres études.

Motif : une méta-analyse d'études d'observation ne vaut pas une méta-analyse d'essais rigoureux.

Il faut donc apprécier séparément la **crédibilité du procédé de revue** et la **certitude des données** réunies.

La revue ne crée pas de preuve : elle en organise la lecture.

La place d'une étude dans la hiérarchie des preuves n'est plus un point de départ ni un point d'arrivée.

Sources : Murad et al., 2016

Un essai randomisé ne donne pas automatiquement une certitude élevée, et une cohorte n'est pas condamnée à une certitude faible.

Le plan d'étude fixe le point de départ. Cinq domaines peuvent le faire descendre, trois peuvent le faire monter. Ce que l'on qualifie au bout du compte, c'est la confiance que l'on peut avoir dans un résultat précis.

Sources : Guyatt et al., 2008 ; Schünemann et al., 2024 ; Murad et al., 2016

4

Ce que la pyramide ne couvre pas

Les études qui ne cherchent pas à démontrer la même chose

Le principe : à chaque question son plan, et son échelle

CE QUE LA PYRAMIDE NE COUVRE PAS

Le constat

Les hiérarchies existantes ont été construites pour une question et une seule : cette intervention produit-elle l'effet attendu ?

Appliquées à une autre question, elles produisent un verdict qui n'a pas de sens.

L'argument

Différents types de questions appellent différents plans d'étude, et donc différentes manières d'apprécier la qualité.

On ne juge pas un cheval de course sur son aptitude au trait : c'est le sens de la formule « horses for courses ».

La conséquence pratique

Une étude ne doit pas être écartée parce qu'elle occupe un rang bas d'une échelle qui ne lui était pas destinée.

La bonne question est : cette étude était-elle le plan qui convenait, et a-t-elle été bien conduite ?

C'est la position à tenir en lecture critique.

Le classement n'est pas un préalable à la lecture : il en est le résultat, et il dépend de la question que l'on pose à l'article.

Sources : Petticrew & Roberts, 2003

Premier cas (1) : les études quasi expérimentales dans les échelles

CE QUE LA PYRAMIDE NE COUVRE PAS

Elles posent bien une question d'efficacité, et elles entrent donc dans la pyramide. Mais leur place y est mal comprise, et l'outil d'appréciation n'est pas le même.

Ce que les échelles en font

La HAS les classe au **niveau 2** si elles sont bien menées : études comparatives non randomisées.

Le JBI leur accorde un **niveau 2** propre, entre l'expérimental et l'observationnel.

Sous GRADE, elles partent d'un degré inférieur, mais peuvent **monter** pour les trois raisons vues plus haut.

Aucune de ces échelles ne les traite comme des essais ratés.

Sources : Haute Autorité de santé, 2013 ; Joanna Briggs Institute, 2014 ; Schünemann et al., 2024

Premier cas (2) : ce qu'il faut demander à une étude quasi expérimentale

CE QUE LA PYRAMIDE NE COUVRE PAS

Ce qu'il faut leur demander

Que les **facteurs de confusion plausibles** soient nommés, et le traitement qui leur a été appliqué décrit.

Que le **groupe de comparaison** soit justifié, et sa non-équivalence discutée.

Que le risque de biais soit apprécié avec ROBINS-I (Risk Of Bias In Non-randomized Studies of Interventions), conçu pour les interventions non randomisées, et non avec RoB 2 (Risk of Bias 2), réservé aux essais randomisés.

Que soit dit **pourquoi** la randomisation n'a pas été employée.

*Sur les interventions organisationnelles en soins, un plan quasi expérimental est souvent le seul possible.
Le disqualifier revient à renoncer à évaluer une grande part de ce que font les infirmiers.*

Sources : Sterne et al., 2016 ; Sterne et al., 2019

Deuxième cas (1) : les études qualitatives et leur échelle

CE QUE LA PYRAMIDE NE COUVRE PAS

Pourquoi la pyramide ne s'applique pas

Une étude qualitative ne cherche pas à établir qu'un effet est réel, ni à en mesurer l'ampleur. Elle cherche à comprendre un vécu, un processus, un milieu. Lui demander un niveau de preuve d'efficacité revient à lui demander ce qu'elle n'a jamais promis.

L'échelle qui lui correspond

Le JBI la classe sur la hiérarchie « signification », où le premier rang revient à une synthèse qualitative.

Pour une synthèse de données qualitatives, l'outil international est GRADE-CERQual, construit sur le modèle de GRADE mais avec ses domaines propres.

Sources : Joanna Briggs Institute, 2014 ; Lewin et al., 2015

Deuxième cas (2) : ce que l'on apprécie dans une étude qualitative

CE QUE LA PYRAMIDE NE COUVRE PAS

Ce que l'on apprécie alors

Les critères de rigueur qualitative : crédibilité, transférabilité, fiabilité, confirmabilité (détaillés dans le cours E1-N2-05 Principes et diversité des méthodes qualitatives).

La rédaction relève de COREQ (Consolidated Criteria for Reporting Qualitative Research) ou de **SRQR** (Standards for Reporting Qualitative Research), qui disent ce qu'un compte rendu doit contenir, sans noter la qualité de l'étude.

Une revue systématique peut porter sur des travaux qualitatifs.

Son résultat n'est pas un effet mesuré, mais un ensemble de constats, dont on apprécie la confiance avec GRADE-CERQual selon quatre composantes : limites méthodologiques, cohérence, adéquation des données, pertinence.

Sources : Lewin et al., 2015 ; Tong et al., 2007 ; O'Brien et al., 2014

Troisième cas (1) : les méthodes mixtes

CE QUE LA PYRAMIDE NE COUVRE PAS

Ces deux champs produisent des connaissances dont les décideurs ont besoin, et qu'aucune pyramide ne classe.

Les méthodes mixtes

Elles combinent une composante quantitative et une composante qualitative, et les font dialoguer.

Aucune échelle de preuve ne leur attribue de rang, parce qu'elles répondent à **plusieurs questions à la fois**.

L'outil de référence est le MMAT (Mixed Methods Appraisal Tool), qui apprécie chaque composante puis leur intégration.

La rédaction se rapporte à GRAMMS (Good Reporting of A Mixed Methods Study).

Sources : Hong et al., 2018 ; O'Cathain et al., 2008

Troisième cas (2) : la recherche en implémentation

CE QUE LA PYRAMIDE NE COUVRE PAS

La recherche en implémentation

Elle étudie le passage d'une intervention efficace à son emploi réel.

Ses résultats sont l'**acceptabilité**, l'**adoption**, la **faisabilité**, la **fidélité**, le **coût**, la **pérennité**.

Aucun de ces résultats n'est un résultat clinique : les classer sur l'échelle de l'efficacité n'a pas de sens.

Les **essais hybrides** permettent de traiter les deux questions dans une même étude.

La plupart des interventions en soins infirmiers sont des interventions complexes, dont l'effet dépend du contexte autant que du contenu. C'est précisément ce que la pyramide ne sait pas représenter.

Sources : Proctor et al., 2011 ; Curran et al., 2012 ; Skivington et al., 2021

Quel outil pour quelle question

CE QUE LA PYRAMIDE NE COUVRE PAS

La question posée	L'outil de classement
Cette intervention est-elle efficace ?	GRADE, ou gradation HAS pour les recommandations françaises
Efficace, sans randomisation possible ?	GRADE, en partant plus bas ; ROBINS-I pour le risque de biais
Comment les personnes vivent-elles cette situation ?	Hiérarchie « signification » du JBI ; GRADE-CERQual pour une synthèse
Comment se combinent chiffres et vécu ?	MMAT ; aucune échelle de niveau ne s'applique
Comment faire adopter une pratique établie ?	Résultats d'implémentation ; aucune échelle de niveau ne s'applique

Deux lignes de ce tableau disent « aucune échelle ne s'applique ». Ce n'est pas un manque du tableau : c'est un fait, et le savoir vous évitera de disqualifier des travaux utiles.

Sources : Schünemann et al., 2024 ; Lewin et al., 2015 ; Hong et al., 2018 ; Proctor et al., 2011

Exercice. Quatre situations de lecture

CE QUE LA PYRAMIDE NE COUVRE PAS

La situation	Que répondez-vous ?
Un collègue écarte une étude parce qu'elle est « seulement qualitative ».	...
Une revue conclut à une certitude faible alors qu'elle ne réunit que des essais randomisés.	...
Une étude d'implémentation ne montre aucun effet sur la mortalité.	...
Une recommandation porte un grade C, mais tout le monde l'applique.	...

Pour chacune, dites quelle confusion est commise, et ce que vous répondez en une phrase.

Sources : Petticrew & Roberts, 2003 ; Haute Autorité de santé, 2013 (exercice construit pour ce cours)

Corrigé. Les quatre situations

CE QUE LA PYRAMIDE NE COUVRE PAS

La situation	La réponse attendue
« Seulement qualitative »	La question n’était pas une question d’efficacité. L’échelle qui convient est la hiérarchie « signification » du JBI, ou CERQual pour une synthèse.
Certitude faible malgré des essais	C’est possible et fréquent : un des cinq domaines a fait descendre le corpus. Cherchez lequel dans les notes du tableau de synthèse (Summary of findings).
Pas d’effet sur la mortalité	Ce n’était pas son but . Une étude d’implémentation mesure l’adoption, la fidélité, la pérennité, non un résultat clinique.
Grade C largement appliqué	Le grade porte sur la recommandation, le niveau de preuve sur les études qui l’appuient. Un accord professionnel solide peut soutenir une pratique peu documentée.

Les quatre erreurs ont la même origine : appliquer à une étude une échelle construite pour une autre question.

Sources : Petticrew & Roberts, 2003 ; Haute Autorité de santé, 2013 ; Schünemann et al., 2024 (corrigé construit pour ce cours)

La pyramide vaut pour les questions d'efficacité. En dehors, elle ne mesure rien, et l'employer quand même produit des jugements faux.

Études quasi expérimentales, qualitatives, en méthodes mixtes, en implémentation : elles ne cherchent pas à démontrer la même chose, et elles ont chacune leur outil d'appréciation. Le savoir fait partie de la lecture critique.

Sources : Petticrew & Roberts, 2003 ; Joanna Briggs Institute, 2014 ; Lewin et al., 2015

Références bibliographiques (1 sur 4)

- American Psychological Association. (2020). *Publication manual of the American Psychological Association* (7e éd.). <https://doi.org/10.1037/0000165-000>
- ANAES. (2000). *Guide d'analyse de la littérature et gradation des recommandations*. Agence nationale d'accréditation et d'évaluation en santé. https://www.has-sante.fr/jcms/c_434715/
- Arrêté du 20 février 2026 relatif au diplôme d'État d'infirmier, NOR ESRS2601184A. (2026). *Journal officiel de la République française*, n° 0048, 25 février 2026. <https://www.legifrance.gouv.fr/jorf/id/JORFTEXT000053568913>
- Concato, J., Shah, N., & Horwitz, R. I. (2000). Randomized, controlled trials, observational studies, and the hierarchy of research designs. *New England Journal of Medicine*, 342(25), 1887-1892. <https://doi.org/10.1056/NEJM200006223422507>
- Curran, G. M., Bauer, M., Mittman, B., Pyne, J. M., & Stetler, C. (2012). Effectiveness-implementation hybrid designs: Combining elements of clinical effectiveness and implementation research to enhance public health impact. *Medical Care*, 50(3), 217-226. <https://doi.org/10.1097/MLR.0b013e3182408812>
- Guyatt, G. H., Oxman, A. D., Vist, G. E., Kunz, R., Falck-Ytter, Y., Alonso-Coello, P., & Schünemann, H. J. (2008). GRADE: An emerging consensus on rating quality of evidence and strength of recommendations. *BMJ*, 336(7650), 924-926. <https://doi.org/10.1136/bmj.39489.470347.AD>

Références bibliographiques (2 sur 4)

Haute Autorité de santé. (2013). *Niveau de preuve et gradation des recommandations de bonne pratique : état des lieux*. https://www.has-sante.fr/jcms/c_1600564

Hong, Q. N., Fàbregues, S., Bartlett, G., Boardman, F., Cargo, M., Dagenais, P., Gagnon, M.-P., Griffiths, F., Nicolau, B., O’Cathain, A., Rousseau, M.-C., Vedel, I., & Pluye, P. (2018). The Mixed Methods Appraisal Tool (MMAT) version 2018 for information professionals and researchers. *Education for Information*, 34(4), 285-291. <https://doi.org/10.3233/EFI-180221>

Joanna Briggs Institute. (2014). *JB I levels of evidence*. Joanna Briggs Institute. <https://jbi.global>

Lewin, S., Glenton, C., Munthe-Kaas, H., Carlsen, B., Colvin, C. J., Gülmezoglu, M., Noyes, J., Booth, A., Garside, R., & Rashidian, A. (2015). Using qualitative evidence in decision making for health and social interventions: An approach to assess confidence in findings from qualitative evidence syntheses (GRADE-CERQual). *PLOS Medicine*, 12(10), e1001895. <https://doi.org/10.1371/journal.pmed.1001895>

Murad, M. H., Asi, N., Alsawas, M., & Alahdab, F. (2016). New evidence pyramid. *Evidence-Based Medicine*, 21(4), 125-127. <https://doi.org/10.1136/ebmed-2016-110401>

O’Brien, B. C., Harris, I. B., Beckman, T. J., Reed, D. A., & Cook, D. A. (2014). Standards for reporting qualitative research: A synthesis of recommendations. *Academic Medicine*, 89(9), 1245-1251. <https://doi.org/10.1097/ACM.0000000000000388>

Références bibliographiques (3 sur 4)

O’Cathain, A., Murphy, E., & Nicholl, J. (2008). The quality of mixed methods studies in health services research. *Journal of Health Services Research & Policy*, 13(2), 92-98. <https://doi.org/10.1258/jhsrp.2007.007074>

OCEBM Levels of Evidence Working Group. (2011). *The Oxford 2011 levels of evidence*. Oxford Centre for Evidence-Based Medicine. <https://www.cebm.ox.ac.uk/resources/levels-of-evidence/ocebmllevels-of-evidence>

Petticrew, M., & Roberts, H. (2003). Evidence, hierarchies, and typologies: Horses for courses. *Journal of Epidemiology and Community Health*, 57(7), 527-529. <https://doi.org/10.1136/jech.57.7.527>

Proctor, E., Silmere, H., Raghavan, R., Hovmand, P., Aarons, G., Bunger, A., Griffey, R., & Hensley, M. (2011). Outcomes for implementation research: Conceptual distinctions, measurement challenges, and research agenda. *Administration and Policy in Mental Health and Mental Health Services Research*, 38(2), 65-76. <https://doi.org/10.1007/s10488-010-0319-7>

Sackett, D. L., Rosenberg, W. M. C., Gray, J. A. M., Haynes, R. B., & Richardson, W. S. (1996). Evidence based medicine: What it is and what it isn’t. *BMJ*, 312(7023), 71-72. <https://doi.org/10.1136/bmj.312.7023.71>

Schünemann, H., Brożek, J., Guyatt, G., & Oxman, A. (dir.). (2013). *GRADE handbook for grading quality of evidence and strength of recommendations*. The GRADE Working Group. <https://gdt.gradepro.org/app/handbook/handbook.html>



Références bibliographiques (4 sur 4)

Schünemann, H. J., Higgins, J. P. T., Vist, G. E., Glasziou, P., Akl, E. A., Skoetz, N., & Guyatt, G. H. (2024). Chapitre 14 : Completing « Summary of findings » tables and grading the certainty of the evidence. Dans J. P. T. Higgins, J. Thomas, J. Chandler, M. Cumpston, T. Li, M. J. Page & V. A. Welch (dir.), *Cochrane handbook for systematic reviews of interventions* (version 6.5). Cochrane. <https://www.training.cochrane.org/handbook>

Skivington, K., Matthews, L., Simpson, S. A., Craig, P., Baird, J., Blazeby, J. M., Boyd, K. A., Craig, N., French, D. P., McIntosh, E., Petticrew, M., Rycroft-Malone, J., White, M., & Moore, L. (2021). A new framework for developing and evaluating complex interventions: Update of Medical Research Council guidance. *BMJ*, 374, n2061. <https://doi.org/10.1136/bmj.n2061>

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

Sterne, J. A. C., Savović, J., Page, M. J., Elbers, R. G., Blencowe, N. S., Boutron, I., Cates, C. J., Cheng, H., Corbett, M. S., Eldridge, S. M., Emberson, J. R., Hernán, M. A., Hopewell, S., Hróbjartsson, A., Junqueira, D. R., Jüni, P., Kirkham, J. J., Lasserson, T., Li, T., ... Higgins, J. P. T. (2019). RoB 2: A revised tool for assessing risk of bias in randomised trials. *BMJ*, 366, l4898. <https://doi.org/10.1136/bmj.l4898>

Tong, A., Sainsbury, P., & Craig, J. (2007). Consolidated criteria for reporting qualitative research (COREQ): A 32-item checklist for interviews and focus groups. *International Journal for Quality in Health Care*, 19(6), 349-357. <https://doi.org/10.1093/intqhc/mzm042>

Mentions légales

CC BY-NC-ND 4.0

Attribution · Pas d'Utilisation Commerciale · Pas de Modification 4.0 International

© Elise Verot, 2026.

Ce cours est mis à disposition selon les termes de la Licence Creative Commons Attribution - Pas d'Utilisation Commerciale - Pas de Modification 4.0 International (CC BY-NC-ND 4.0).

Ce document peut être copié et partagé dans sa forme originale, sous réserve de l'attribution de l'auteur et en dehors de tout usage commercial. Toute version adaptée, transformée ou modifiée ne peut pas être diffusée dans le cadre de la licence CC BY-NC-ND 4.0. Toute réutilisation ou diffusion autorisée doit obligatoirement mentionner explicitement le nom de l'auteur original.

Texte intégral de la licence : <https://creativecommons.org/licenses/by-nc-nd/4.0/deed.fr>

Transparence sur l'usage de l'intelligence artificielle. Les supports pédagogiques de l'UE E.1 ont été conçus, rédigés et sourcés par l'auteure. Une intelligence artificielle générative conversationnelle a été utilisée sous son contrôle pour l'harmonisation de la charte graphique et de la mise en page de l'ensemble des cours, la mise aux normes APA 7 des références et la vérification de leurs DOI, ainsi que le contrôle de cohérence et de conformité entre les documents. Toute correction a été validée par l'auteure, qui assume la responsabilité du contenu.

Merci de votre attention

Elise VEROT

Professeure des Universités en Sciences Infirmières
Faculté de Santé, Université Jean Monnet, Saint-Étienne
CHU de Saint-Étienne · UR 4129 P2S · CIC 1408 INSERM

elise.verot@univ-st-etienne.fr

Vous savez désormais lire un niveau de preuve, un grade et un degré de certitude, et dire à quelle question chacun répond. Vous savez aussi reconnaître les travaux qu'aucune échelle ne classe, et pourquoi.