

Lire et critiquer un article scientifique

Séance 3 sur 3 : Les limites méthodologiques

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



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

Cette séance dans le cours complet

LES 3 SÉANCES DU COURS

- Séance 1. Anatomie d'un article scientifique
- Séance 2. Grilles de lecture et résultat principal
- **Séance 3. Les limites méthodologiques ← vous êtes ici**

Cours E1-N2-04 Lire et critiquer un article scientifique. Séance 3 sur 3.

Objectifs du cours

CE QUE VOUS SAUREZ FAIRE À L'ISSUE DES TROIS SÉANCES

- **Reconnaître** la structure **IMRaD** d'un article et les éléments périphériques qui l'accompagnent, et dire ce que chaque section doit contenir ;
- **Lire** un article en trois passes plutôt que dans l'ordre du texte ;
- **Identifier** la ligne directrice de rapport qui s'applique à un type d'étude donné, et s'en servir pour vérifier ce qui manque ;
- **Choisir** une **grille de lecture** adaptée, et distinguer une grille de lecture d'un outil d'évaluation du **risque de biais** ;
- **Repérer** le **critère de jugement** principal d'une étude et le résultat qui s'y rapporte ;
- **Interpréter** un **intervalle de confiance**, une **valeur p**, un **risque relatif**, une différence de risque et un nombre de sujets à traiter ; lire un **forest plot** et un graphique en entonnoir ;
- **Distinguer** erreur aléatoire et erreur systématique, et nommer les quatre **validités** ;
- **Reconnaître** les **biais** de sélection, d'information et de confusion les plus fréquents en soins, et distinguer une association d'une relation causale ;
- **Apprécier** la rigueur d'une étude qualitative au moyen de ses quatre critères ;
- **Rédiger** la section « limites » de son propre travail.

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

4

Les limites méthodologiques

Séance 3 — Repérer les biais, mesurer sa confiance



Quatre questions de bon sens, avant toute technique

Support de médiation scientifique : la bande dessinée « Comment savoir si une étude scientifique est fiable ? » du site Curieux !, média de culture scientifique soutenu par le ministère de l'Enseignement supérieur et de la Recherche.

Accès au support en ligne : Joumard, M. (2025, 22 janvier). Comment savoir si une étude scientifique est fiable ? [Bande dessinée]. Curieux! <https://www.curieux.live/2025/01/22/bd-comment-savoir-si-une-etude-est-fiable/>

- D'où vient cette étude ? Qui l'a produite, qui l'a financée, où a-t-elle été publiée ?
- Quelle méthodologie ? Quelle taille et quelle diversité d'échantillon, quel comparateur, quelle durée ?
- Quelle réception par les pairs ? L'étude a-t-elle été relue, commentée, reprise, contestée ?
- La science évolue : il est normal que de nouveaux travaux remettent en cause les précédents.

Ces quatre questions ne remplacent pas une analyse méthodologique. Elles la précèdent, et elles suffisent souvent à écarter ce qui ne mérite pas de lecture approfondie.

Point de vigilance : une ressource de médiation ne se cite pas comme une preuve.

Sources : Joumard, 2025

Erreur aléatoire et erreur systématique

Distinction fondatrice

Erreur aléatoire

Le hasard de l'échantillonnage

- Elle vient du fait qu'on observe un échantillon et non la population entière.
- Elle est sans direction : elle peut faire surestimer comme sous-estimer.
- Elle diminue quand l'effectif augmente.
- **Elle se quantifie** : c'est l'objet de l'intervalle de confiance et de la valeur p.
- Elle affecte la précision.

Erreur systématique — le biais

Un défaut du protocole

- Elle vient de la façon dont l'étude a été conçue, conduite ou analysée.
- Elle a une direction : elle pousse le résultat toujours du même côté.
- Elle ne diminue pas quand l'effectif augmente : une grande étude biaisée est une grande erreur.
- Elle ne se quantifie pas par un test statistique. Elle se raisonne.
- Elle affecte la justesse.

Point de vigilance : c'est pourquoi un très grand essai mal conçu peut être moins fiable qu'un "petit" essai rigoureux : la statistique corrige le hasard, jamais le biais.

Sources : Shadish et al., 2002 ; Catalogue of Bias Collaboration, s. d.

Quatre validités à distinguer

Cadre

- **Validité interne** : l'effet observé est-il bien dû à ce que l'on croit ? C'est la question des biais et des facteurs de confusion. Sans elle, rien d'autre n'a d'importance.
- **Validité externe** : le résultat est-il transposable à d'autres personnes, d'autres lieux, d'autres moments ? On parle aussi de généralisabilité ou, en recherche qualitative, de transférabilité.
- **Validité de construit** : l'instrument mesure-t-il réellement ce qu'il prétend mesurer ? Une échelle d'anxiété mesure-t-elle l'anxiété, ou l'inconfort physique ?
- **Validité de conclusion statistique** : les analyses conduites permettent-elles les conclusions tirées ? Puissance suffisante, tests appropriés, absence de multiplication non contrôlée des comparaisons.

Point de vigilance : elles sont hiérarchisées : une étude sans validité interne ne peut pas avoir de validité externe. Un résultat faux n'est transposable nulle part.

Sources : Shadish et al., 2002

Les trois grandes familles de biais

Taxonomie

Trois familles, trois moments : qui entre dans l'étude, comment on mesure, ce qui se cache derrière l'association.

- **Biais de sélection** : les personnes étudiées ne représentent pas celles auxquelles on veut appliquer le résultat, ou les groupes comparés ne sont pas comparables. Survient au recrutement, à la constitution des groupes, ou à la sortie de l'étude.
- **Biais d'information** : les données recueillies ne reflètent pas fidèlement la réalité. Survient à la mesure : instrument, personne qui mesure, personne qui répond.
- **Facteur de confusion** : un troisième facteur, lié à la fois à l'exposition et à l'événement, brouille la relation observée. Ce n'est pas une erreur de recueil : c'est une structure de la réalité qu'il faut neutraliser.

Point de vigilance : le Catalogue of Bias (catalogofbias.org), maintenu par le Centre for Evidence-Based Medicine d'Oxford, recense et documente les biais un par un, avec exemples et méthodes de réduction.

Sources : Catalogue of Bias Collaboration, s. d.

Les biais de sélection les plus fréquents en soins

SÉLECTION

Biais	Mécanisme	Exemple en sciences infirmières
Biais de volontariat	Ceux qui acceptent de participer diffèrent de ceux qui refusent.	Une enquête sur l'épuisement professionnel à laquelle répondent surtout les soignants les plus engagés, ou les plus en souffrance.
Biais de non-réponse	Les non-répondants diffèrent systématiquement des répondants.	Un questionnaire adressé par courriel dont le taux de réponse est de 25 % : que pensent les 75 % restants ?
Biais d'attrition	Les personnes perdues de vue ne sont pas un sous-groupe au hasard.	Dans un essai d'éducation thérapeutique, ceux qui abandonnent sont souvent ceux pour qui l'intervention fonctionne le moins.
Biais d'indication	L'indication du soin est liée à la gravité, donc au pronostic. Ce mécanisme, souvent classé comme une confusion par indication, apparaît dès la constitution des groupes.	Les patients recevant un matelas à air ont plus d'escarres, parce qu'on le prescrit aux plus à risque, non parce qu'il en cause.
Biais du survivant	On n'observe que ceux qui sont encore là pour être observés.	Étudier la satisfaction des soignants d'un service en n'interrogeant que ceux qui n'en sont pas partis.
Biais de recrutement en centre spécialisé	La population d'un centre de référence n'est pas celle du terrain.	Des résultats obtenus dans un CHU de recours transposés à un service de proximité.

Sources : *Catalogue of Bias Collaboration, s. d.*

Les biais d'information les plus fréquents en soins

INFORMATION

Biais	Mécanisme	Réduction possible
Biais de mémorisation	Les personnes ayant vécu un événement s'en souviennent mieux et le relient plus facilement à une cause.	Recueil prospectif ; données objectives issues des dossiers plutôt que déclaratives.
Biais de désirabilité sociale	On répond ce qu'on pense être attendu, surtout sur les pratiques professionnelles ou les comportements de santé.	Anonymat réel, auto-questionnaire, formulation neutre des questions.
Effet Hawthorne	Se savoir observé modifie le comportement, indépendamment de l'intervention.	Période d'observation préalable, données recueillies en routine, groupe témoin également observé.
Biais de l'observateur	Celui qui mesure connaît le groupe d'affectation et interprète en conséquence.	Insu de l'évaluateur : presque toujours possible même quand l'insu des soignants ne l'est pas.
Biais de classement	L'exposition ou l'événement est mal catégorisé.	Définitions explicites et écrites à l'avance, formation des enquêteurs, double codage.
Instrument non validé	Le questionnaire mesure autre chose que ce qu'il annonce, ou pas de manière stable.	Utiliser des échelles validées, et validées dans la langue et la population concernées.

Point de vigilance : l'effet Hawthorne est particulièrement puissant en recherche sur les soins : l'hygiène des mains s'améliore dès que l'on sait qu'un audit est en cours.

Sources : Catalogue of Bias Collaboration, s. d.

© Elise Verot, 2026 – Licence CC BY-NC-ND 4.0. Partage à l'identique uniquement. Modification, transformation et usage commercial strictement interdits. Citer l'auteur.

Le facteur de confusion et sa neutralisation

Confusion

Trois conditions à réunir

- 1. Le facteur est associé à l'exposition étudiée.
- 2. Le facteur est un facteur de risque de l'événement, indépendamment de l'exposition.
- 3. Le facteur n'est pas sur le chemin causal entre l'exposition et l'événement.
- Exemple : on observe que les services où les infirmiers sont les plus expérimentés ont moins d'escarres. L'ancienneté du service, sa spécialité, la lourdeur de ses patients sont des facteurs de confusion candidats.

Comment on le neutralise

- **Avant l'étude** : randomisation : la seule méthode qui traite aussi les facteurs inconnus.
- **Avant l'étude** : restriction : n'inclure qu'une catégorie de patients. Gain de validité interne, perte de validité externe.
- **Avant l'étude** : appariement : associer à chaque cas un témoin de mêmes caractéristiques.
- **Après l'étude** : stratification, ajustement dans un modèle multivarié, score de propension.
- Toutes les méthodes « après » ont la même limite : elles ne corrigent que ce qui a été mesuré.

Point de vigilance : quand un article annonce un résultat « ajusté », cherchez la liste des variables d'ajustement. Ce qui n'y figure pas n'a pas été corrigé.

Sources : Catalogue of Bias Collaboration, s. d. ; Sterne et al., 2016

Quatre pièges de l'analyse

Statistique

- **Puissance insuffisante** : la probabilité de détecter un effet réel dépend de l'effectif et de la taille de l'effet recherché. Une étude sous-dimensionnée conclura à l'absence d'effet alors qu'elle n'a simplement pas pu le voir. À vérifier : le calcul d'effectif est-il présenté et justifié ?
- **Multiplicité des tests** : plus on teste, plus on trouve. Avec un seuil à 5 %, tester vingt critères produit en moyenne un résultat « significatif » par pur hasard. À vérifier : combien de comparaisons ont été faites, et le seuil a-t-il été ajusté ?
- **Analyses en sous-groupes** : un effet trouvé dans un sous-groupe non prévu à l'avance est une hypothèse, pas un résultat. Elles doivent être annoncées dans le protocole et présentées avec un test d'interaction.
- **Formulation d'hypothèses après coup** : présenter comme prévu une hypothèse formulée après avoir vu les données. L'enregistrement préalable du protocole est la parade principale.

Point de vigilance : Ioannidis (2005) a montré comment la combinaison de faible puissance, de nombreuses hypothèses testées et de flexibilité analytique rend une part importante des résultats publiés difficilement reproductible.

Sources : Ioannidis, 2005 ; Hopewell et al., 2025

Quatre critères de rigueur (et ce qui les démontre)

Recherche qualitative

Critères proposés par Lincoln et Guba (1985). Les grilles CASP « qualitative » et COREQ en recouvrent l'esprit (réflexivité, description du contexte, traçabilité de l'analyse) sans reprendre cette terminologie.

- **Crédibilité** : les résultats représentent-ils fidèlement l'expérience des participants ? Ce qui la démontre : triangulation des sources ou des chercheurs, validation par les participants, présence d'extraits d'entretiens illustrant chaque thème.
- **Transférabilité** : les résultats peuvent-ils éclairer un autre contexte ? Ce qui la démontre : une description dense du contexte et des participants, suffisante pour que le lecteur juge lui-même de la transposition.
- **Fiabilité (dependability)** : la démarche est-elle traçable et cohérente ? Ce qui la démontre : un journal de recherche, une description explicite du processus d'analyse, un audit externe.
- **Confirmabilité** : les résultats découlent-ils des données plutôt que des présupposés du chercheur ? Ce qui la démontre : la réflexivité, par laquelle le chercheur explicite sa position, son rapport au terrain et son influence possible.

Point de vigilance : reprocher à une étude qualitative son faible effectif, son absence de randomisation ou l'absence de valeur p est une erreur de cadre.

Sources : Lincoln & Guba, 1985 ; Critical Appraisal Skills Programme, s. d. ; Tong et al., 2007

Ce qui est une limite, ce qui n'en est pas

Discernement

De vraies limites

- Un critère de jugement principal changé en cours de route.
- Une perte de suivi importante et non expliquée.
- Une absence d'insu de l'évaluateur alors qu'il était réalisable.
- Un instrument de mesure non validé dans la population étudiée.
- Une conclusion qui excède les données (le spin).
- Un conflit d'intérêts non déclaré.

Ce qu'on prend à tort pour des limites

Critiques fréquentes et infondées

- **Un petit effectif en recherche qualitative** : ce n'est pas le bon critère.
- **Un résultat négatif** : un essai bien conduit qui ne trouve rien produit une information utile.
- **Un caractère monocentrique assumé et décrit** : c'est une limite de portée, pas un défaut de validité.
- **L'absence d'insu des soignants quand elle était matériellement impossible.**
- Une revue publiée dans une revue à faible facteur d'impact.

Point de vigilance : la qualité d'une critique se mesure à sa capacité à indiquer la direction du biais : ce défaut fait-il surestimer ou sous-estimer l'effet ?

Sources : Boutron et al., 2010 ; Sterne et al., 2019

Association et causalité : ce qui permet d'aller plus loin

Causalité

Bradford Hill (1965) a proposé un faisceau de neuf points de vue, et non un algorithme, pour discuter une causalité à partir de données observationnelles. Les six retenus ici, plausibilité et cohérence étant réunies, sont les plus utiles en lecture d'article ; spécificité et analogie, rarement décisives, sont laissées de côté.

- **Force de l'association** : plus l'effet est important, moins il est facilement explicable par un biais résiduel.
- **Constance** : le même résultat est retrouvé dans des populations et des contextes différents.
- **Temporalité** : la cause précède l'effet. C'est le seul argument indispensable : les autres ne sont que des appuis.
- **Gradient biologique** : plus d'exposition, plus d'effet.
- **Plausibilité et cohérence** : un mécanisme explicatif compatible avec les connaissances existantes.
- **Preuve expérimentale** : le retrait de l'exposition fait disparaître l'effet.

Ces arguments ne se cochent pas comme une liste : ils s'assemblent en raisonnement, et Hill lui-même mettait en garde contre tout usage mécanique.

Sources : Hill, 1965

Conflits d'intérêts et intégrité scientifique

Intégrité

Ce que vous devez chercher dans l'article

- **Qui a financé l'étude et à quelle hauteur.** Si le financeur est intéressé au résultat, a-t-il eu accès aux données, à l'analyse, à la rédaction ? Les auteurs déclarent-ils des liens d'intérêt ?
- **Un lien d'intérêt déclaré n'invalide pas un travail.** Un lien non déclaré est un problème d'intégrité.

Les quatre principes de l'intégrité scientifique

- **Fiabilité** : garantir la qualité de la conception, de la méthodologie, de l'analyse et de l'usage des ressources.
- **Honnêteté** : mener et rapporter la recherche de manière transparente, juste, complète et objective.
- **Respect** : envers les collègues, les participants, la société et l'environnement.
- **Responsabilité** : de l'idée à la publication, y compris la formation et l'encadrement.

Point de vigilance : ces principes sont ceux du Code de conduite européen pour l'intégrité en recherche (ALLEA). En France, la loi de programmation de la recherche de 2020 a inscrit l'intégrité scientifique dans le code de la recherche.

Le cas particulier des interventions complexes

Sciences infirmières

Ce qui les caractérise

- **Plusieurs composants qui interagissent entre eux.**
- **Un effet dépendant du contexte** : équipe, organisation, relation, culture de service.
- **Une mise en œuvre variable** d'un lieu à l'autre, et même d'un professionnel à l'autre.
- **Conséquence** : un résultat négatif peut signifier « l'intervention ne marche pas » ou « l'intervention n'a pas été mise en œuvre comme prévu ». Ce ne sont pas les mêmes conclusions.

Ce qu'une bonne étude ajoute

- **Une évaluation de processus** : l'intervention a-t-elle été effectivement délivrée, à quelle dose, avec quelle fidélité ?
- Une description complète de l'intervention selon TIDieR.
- **Une théorie du changement explicite** : par quel mécanisme l'intervention est-elle censée produire son effet ?
- **Souvent, un volet qualitatif imbriqué pour comprendre les obstacles et les leviers de mise en œuvre.**

Cadre de référence : Skivington et al. (2021), mise à jour des recommandations du Medical Research Council.

Point de vigilance : l'essai 6-PACK illustre exactement ce point : le programme a bien été appliqué et n'a pourtant pas réduit les chutes.

Sources : Skivington et al., 2021 ; Hoffmann et al., 2014 ; Barker et al., 2016

Rédiger la section « limites » de votre propre travail

Production

Attendu de l'évaluation de l'UE E.1 : une synthèse structurée comportant une analyse critique argumentée.

- **1. Identifier : nommer la limite avec son terme technique** : biais de sélection, biais de mesure, facteur de confusion résiduel, puissance insuffisante.
- **2. Localiser : dire où elle se situe dans l'étude** : recrutement, mesure, analyse, interprétation.
- **3. Orienter : indiquer la direction probable** : ce défaut conduit-il à surestimer ou à sous-estimer l'effet ?
- **4. Pondérer : hiérarchiser** : cette limite remet-elle en cause la conclusion, ou en réduit-elle seulement la portée ?
- **5. Conclure : formuler ce que l'étude permet de dire, et ce qu'elle ne permet pas de dire**. C'est la phrase que votre jury attend.

Point de vigilance : une liste de limites sans direction ni hiérarchie ne démontre pas une compétence critique : elle démontre qu'on connaît le vocabulaire.

Sources : Arrêté du 20 février 2026

Ce qui rend une recherche vérifiable

Science ouverte

- **Enregistrement préalable** : le protocole et le critère de jugement principal sont déposés avant le début de l'étude. Registres : ClinicalTrials.gov pour les essais, PROSPERO pour les revues systématiques. Ce dépôt permet de vérifier que ce qui est publié correspond à ce qui était prévu.
- **Accès ouvert** : la publication est librement consultable. Beaucoup des références de ce cours le sont, et c'est ce qui rend une lecture critique possible sans abonnement institutionnel.
- **Partage des données et du code** : les données et les scripts d'analyse sont déposés, quand la protection des personnes le permet. Une déclaration de disponibilité des données figure désormais dans la plupart des revues.
- **Réplication** : un résultat n'acquiert sa solidité qu'en étant retrouvé par d'autres équipes, dans d'autres contextes. C'est la constance au sens de Bradford Hill.

Point de vigilance : protection des données : la recherche en santé traite des données personnelles. Anonymisation, base légale, information des personnes et déclaration relèvent du RGPD et conditionnent la licéité de l'étude.

Sources : Hopewell et al., 2025 ; Page et al., 2021

Exercice d'application — séance 3

Travail sur article

Le support

- **Aiken et al. (2012)** : enquêtes transversales auprès d'infirmiers et de patients dans 12 pays d'Europe et aux États-Unis (BMJ).
- Article en accès libre.
- Référence complète en diapositive 29 de ce diaporama.

Le travail demandé

- 1. Identifier trois limites méthodologiques et, pour chacune : la nommer, la localiser, indiquer la direction probable du biais, en apprécier le poids.
- 2. Discuter ce que le design transversal autorise et interdit de conclure.
- 3. Formuler en deux phrases ce que cette étude permet de dire à un cadre de santé qui vous interroge sur les effectifs.

Point de vigilance : cet exercice prépare directement l'évaluation de l'UE : synthèse structurée, écrite et orale, d'une analyse critique de la littérature.

Sources : Aiken et al., 2012 ; von Elm et al., 2007

Corrigé type de l'exercice 3

CE QUE L'ON ATTEND (AIKEN ET AL., 2012)

Trois limites, avec leur localisation et leur direction

- **1. Biais d'information** : la qualité des soins et l'environnement de travail sont déclarés par les mêmes infirmiers, au même moment ; une perception commune peut renforcer l'association observée (surestimation probable). **2. Biais de sélection par non-réponse** : les taux de réponse varient selon les pays et les répondants peuvent différer des non-répondants ; direction incertaine. **3. Confusion résiduelle** : les ajustements ne corrigent que les caractéristiques mesurées des hôpitaux et des infirmiers ; la lourdeur des patients ou le type d'établissement peuvent expliquer une part de l'association.

Ce que le design transversal autorise, et ce qu'il interdit

- **Il autorise à** décrire (de 11 % à 47 % d'infirmiers jugeant la qualité des soins médiocre ou passable selon le pays) et à repérer des associations : OR ajusté 0,56 (IC 95 % 0,51-0,61) pour un meilleur environnement de travail ; 1,11 (1,07-1,15) par patient supplémentaire par infirmier. **Il interdit d'établir le sens de la relation** : tout est mesuré au même moment, le point de vue de temporalité de Hill n'est pas satisfait.

En deux phrases, pour un cadre de santé

- Dans douze pays, les hôpitaux où la charge en patients par infirmier est plus élevée et l'environnement de travail moins favorable sont aussi ceux où infirmiers et patients rapportent une moins bonne qualité des soins. Cette association, cohérente d'un pays à l'autre, ne prouve pas qu'augmenter les effectifs améliorerait la qualité : elle le rend plausible et justifie de le tester.

Sources : Aiken et al., 2012 ; Hill, 1965 ; von Elm et al., 2007

Vérifications

Auto-évaluation 3

La question

Cas construit, sans lien avec l'article de l'exercice. Une étude transversale menée dans douze hôpitaux montre que les services où les infirmiers déclarent le plus d'épuisement professionnel sont aussi ceux où les patients rapportent la moins bonne qualité de soins. Quelle conclusion est méthodologiquement correcte ?

- A. L'épuisement professionnel dégrade la qualité des soins
- B. La mauvaise qualité des soins provoque l'épuisement professionnel
- C. Il existe une association entre les deux ; le design transversal ne permet pas d'en établir le sens, et un facteur commun (la dotation en personnel, par exemple) peut expliquer les deux
- D. L'association est probablement due au hasard, car les études transversales ne sont pas fiables

La réponse

C. Une étude transversale mesure tout au même moment : le critère de temporalité de Bradford Hill, seul argument indispensable à la causalité, n'est pas satisfait. **Les réponses A et B sont toutes deux plausibles et aucune n'est démontrée.** La troisième possibilité, celle d'un facteur de confusion commun (typiquement le ratio patients/infirmier ou l'environnement de travail), est ici la plus vraisemblable. **La réponse D confond un design inadapté à la causalité avec une absence de fiabilité** : le design transversal est parfaitement valide pour décrire une prévalence et repérer des associations.

Sources : Hill, 1965

De la question clinique à la décision : la démarche complète (1/2)

SYNTHÈSE DU COURS

Étape	Ce que vous faites	Outil mobilisé
1. Formuler	Transformer une interrogation de terrain en question de recherche structurée.	PICO
2. Chercher	Interroger plusieurs bases avec une équation reproductible, et documenter la démarche.	PubMed (MeSH), Cochrane Library, CINAHL, LiSSa, Cairn (cours E1-N1-05 La recherche documentaire en santé)
3. Identifier	Reconnaître le type d'étude et vérifier qu'il est adapté à la question.	Les trois passes et les sept blocs (séance 1)
4. Situer	Positionner la preuve et son degré de certitude.	Cours E1-N2-03 Les niveaux de preuve

De la question clinique à la décision : la démarche complète (2/2)

SYNTHÈSE DU COURS

Étape	Ce que vous faites	Outil mobilisé
5. Analyser	Parcourir l'article de façon reproductible et tracée.	CASP, JBI, grille en dix questions
6. Interpréter	Lire le résultat principal : mesure d'effet, intervalle de confiance, pertinence clinique.	RR, OR, RD, NNT, MD, HR ; forest plot
7. Critiquer	Identifier les limites, leur direction et leur poids.	Taxonomie des biais, RoB 2, ROBINS-I, AMSTAR 2
8. Décider	Confronter la preuve à l'expertise clinique et aux préférences de la personne.	Le modèle de la pratique fondée sur les preuves

Point de vigilance : l'étape 8 est celle qui donne son sens aux sept premières.
Une analyse critique qui ne débouche sur aucune décision est un exercice, pas une compétence professionnelle.

Sources : Critical Appraisal Skills Programme, s. d. ; Sterne et al., 2019 ; Sterne et al., 2016 ; Shea et al., 2017

Séance 3 — À retenir

Synthèse

- **L'erreur aléatoire se corrige par l'effectif ; le biais ne se corrige pas.** Une grande étude biaisée est une grande erreur.
- **Trois familles de biais, trois moments :** qui entre dans l'étude, comment on mesure, ce qui se cache derrière l'association.
- **Les méthodes d'ajustement ne corrigent que ce qui a été mesuré.** Cherchez toujours la liste des variables d'ajustement.
- **La recherche qualitative a ses propres critères de rigueur :** crédibilité, transférabilité, fiabilité, confirmabilité.
- **Une critique de qualité indique la direction du biais et hiérarchise.** Une liste de termes techniques ne suffit pas.
- **Association n'est pas causalité ;** la temporalité est le seul argument indispensable.

Glossaire (1 sur 3)

ALLOCATION DISSIMULÉE : Procédure garantissant que la personne qui inclut un participant ne peut pas connaître à l'avance le bras d'affectation.

ATTRITION : Perte de participants au cours du suivi. Devient un biais lorsque les sortants diffèrent des restants.

BIAIS : Erreur systématique liée à la conception, à la conduite ou à l'analyse d'une étude. Elle a une direction et ne diminue pas avec l'effectif.

CONFUSION (facteur de) : Troisième facteur associé à la fois à l'exposition et à l'événement, sans être sur le chemin causal entre les deux.

CRITÈRE DE JUGEMENT PRINCIPAL : Résultat sur lequel l'étude est dimensionnée et engage sa conclusion. Il doit être unique et annoncé dans les méthodes.

GRADE : Système d'évaluation du niveau de certitude d'un corpus de preuves, critère de jugement par critère de jugement, en quatre niveaux.

HÉTÉROGÉNÉITÉ : Variabilité entre les résultats des études d'une méta-analyse, au-delà de ce que le hasard explique. Quantifiée notamment par I^2 .

IMRaD : Structure conventionnelle d'un article scientifique : Introduction, Méthodes, Résultats, Discussion.

INSU (aveugle) : Fait de tenir un acteur de l'étude dans l'ignorance du groupe d'affectation (groupe "intervention ou expérimental" ou groupe "contrôle": participant, soignant, évaluateur, analyste).

Glossaire (2 sur 3)

INTENTION DE TRAITER : Analyse dans laquelle chaque participant est comptabilisé dans le groupe où il a été randomisé, quelle que soit l'intervention réellement reçue.

INTERVALLE DE CONFIANCE : Plage de valeurs compatibles avec les données observées, sous le modèle statistique utilisé. Sa position et sa largeur s'interprètent ensemble.

MESH : Vocabulaire contrôlé d'indexation de la base MEDLINE, permettant une recherche par concept et non par mot.

NNT (nombre de sujets à traiter) : Nombre de personnes à prendre en charge pour éviter un événement. Inverse de la différence de risque.

ODDS RATIO ("rapport de cotes") : Rapport de cotes entre deux groupes. Proche du risque relatif quand l'événement est rare, s'en éloigne quand il est fréquent.

PER-PROTOCOLE : Analyse restreinte aux participants qui ont effectivement suivi le protocole prévu. Elle rompt l'équilibre créé par la randomisation ; l'intention de traiter reste l'analyse principale.

PICO : Format de structuration d'une question de recherche : population, intervention, comparateur, critère de jugement.

PUISSANCE : Probabilité de détecter un effet réel. Dépend de l'effectif et de la taille de l'effet recherché.

RANDOMISATION : Attribution de l'intervention par tirage au sort. Seule méthode qui équilibre aussi les facteurs de confusion non mesurés.

RANDOMISATION EN GRAPPES (clusters) : Randomisation d'unités entières (services, établissements) plutôt que d'individus.

Glossaire (3 sur 3)

RÉFLEXIVITÉ : En recherche qualitative, explicitation par le chercheur de sa position, de son rapport au terrain et de son influence possible sur les données.

RISQUE RELATIF : Rapport des risques entre deux groupes. À distinguer de la différence de risque, seule porteuse d'un sens clinique direct.

REVUE SYSTÉMATIQUE : Synthèse conduite selon un protocole préétabli, avec recherche exhaustive, sélection reproductible et évaluation du risque de biais.

SATURATION : En recherche qualitative, moment où les données nouvelles n'apportent plus d'élément inédit à l'analyse.

SPIN : Présentation d'un résultat au-delà de ce que les données autorisent, notamment dans le résumé et la conclusion.

VALIDITÉ EXTERNE : Degré auquel un résultat peut être transposé à d'autres personnes, d'autres lieux, d'autres moments.

VALIDITÉ INTERNE : Degré auquel l'effet observé est réellement attribuable à ce que l'étude prétend étudier.

VALEUR P ("petit p" / p-value) : La valeur p indique à quel point le résultat observé est compatible avec l'hypothèse qu'il n'y a pas de différence ou d'effet.

Repère : $p <$ au seuil de significativité fixé dans l'étude → résultat considéré comme statistiquement significatif ; $p >$ ou $=$ à ce seuil → résultat non statistiquement significatif.

À vérifier dans chaque article : quel seuil de significativité les auteurs ont choisi, souvent noté α (alpha). Plus p est petite par rapport au seuil fixé, moins le résultat est compatible avec l'absence de différence ou d'effet. Le seuil est souvent 0,05, mais il peut être différent selon le protocole ou les analyses.

Références bibliographiques (1 sur 4)

Aiken, L. H., Sermeus, W., Van den Heede, K., Sloane, D. M., Busse, R., McKee, M., Bruyneel, L., Rafferty, A. M., Griffiths, P., Moreno-Casbas, M. T., Tishelman, C., Scott, A., Brzostek, T., Kinnunen, J., Schwendimann, R., Heinen, M., Zikos, D., Sjetne, I. S., Smith, H. L., & Kutney-Lee, A. (2012). Patient safety, satisfaction, and quality of hospital care: Cross sectional surveys of nurses and patients in 12 countries in Europe and the United States. *BMJ*, 344, e1717. <https://doi.org/10.1136/bmj.e1717>

All European Academies. (2023). *The European code of conduct for research integrity* (Revised edition 2023). <https://allea.org/code-of-conduct/>

Arrêté du 20 février 2026 relatif au diplôme d'État d'infirmier, NOR ESR52601184A. (2026). *Journal officiel de la République française*, n° 0048, 25 février 2026. <https://www.legifrance.gouv.fr/jorf/id/JORFTEXT000053568913>

Barker, A. L., Morello, R. T., Wolfe, R., Brand, C. A., Haines, T. P., Hill, K. D., Brauer, S. G., Botti, M., Cumming, R. G., Livingston, P. M., Sherrington, C., Zavarsek, S., Lindley, R. I., & Kamar, J. (2016). 6-PACK programme to decrease fall injuries in acute hospitals: Cluster randomised controlled trial. *BMJ*, 352, h6781. <https://doi.org/10.1136/bmj.h6781>

Boutron, I., Dutton, S., Ravaud, P., & Altman, D. G. (2010). Reporting and interpretation of randomized controlled trials with statistically nonsignificant results for primary outcomes. *JAMA*, 303(20), 2058-2064. <https://doi.org/10.1001/jama.2010.651>

Catalogue of Bias Collaboration. (s. d.). *Catalogue of Bias*. Centre for Evidence-Based Medicine, University of Oxford. <https://catalogofbias.org/>



Références bibliographiques (2 sur 4)

Critical Appraisal Skills Programme. (s. d.). *CASP checklists*. <https://casp-uk.net/casp-tools-checklists/>

Hill, A. B. (1965). The environment and disease: Association or causation? *Proceedings of the Royal Society of Medicine*, 58(5), 295-300. <https://doi.org/10.1177/003591576505800503>

Hoffmann, T. C., Glasziou, P. P., Boutron, I., Milne, R., Perera, R., Moher, D., Altman, D. G., Barbour, V., Macdonald, H., Johnston, M., Lamb, S. E., Dixon-Woods, M., McCulloch, P., Wyatt, J. C., Chan, A.-W., & Michie, S. (2014). Better reporting of interventions: Template for intervention description and replication (TIDieR) checklist and guide. *BMJ*, 348, g1687. <https://doi.org/10.1136/bmj.g1687>

Hopewell, S., Chan, A.-W., Collins, G. S., Hróbjartsson, A., Moher, D., Schulz, K. F., Tunn, R., Aggarwal, R., Berkwits, M., Berlin, J. A., Bhandari, N., Butcher, N. J., Campbell, M. K., Chidebe, R. C. W., Elbourne, D., Farmer, A., Fergusson, D. A., Golub, R. M., Goodman, S. N., ... Boutron, I. (2025). CONSORT 2025 statement: Updated guideline for reporting randomised trials. *BMJ*, 389, e081123. <https://doi.org/10.1136/bmj-2024-081123>

Ioannidis, J. P. A. (2005). Why most published research findings are false. *PLoS Medicine*, 2(8), e124. <https://doi.org/10.1371/journal.pmed.0020124>

Joumard, M. (2025, 22 janvier). *Comment savoir si une étude scientifique est fiable ?* [Bande dessinée]. Curieux! <https://www.curieux.live/2025/01/22/bd-comment-savoir-si-une-etude-est-fiable/>

Références bibliographiques (3 sur 4)

Lincoln, Y. S., & Guba, E. G. (1985). *Naturalistic inquiry*. Sage.

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., ... Moher, D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ*, 372, n71. <https://doi.org/10.1136/bmj.n71>

Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Houghton Mifflin.

Shea, B. J., Reeves, B. C., Wells, G., Thuku, M., Hamel, C., Moran, J., Moher, D., Tugwell, P., Welch, V., Kristjansson, E., & Henry, D. A. (2017). AMSTAR 2: A critical appraisal tool for systematic reviews that include randomised or non-randomised studies of healthcare interventions, or both. *BMJ*, 358, j4008. <https://doi.org/10.1136/bmj.j4008>

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>

Références bibliographiques (4 sur 4)

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>

von Elm, E., Altman, D. G., Egger, M., Pocock, S. J., Gøtzsche, P. C., & Vandenbroucke, J. P. (2007). The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: Guidelines for reporting observational studies. *The Lancet*, 370(9596), 1453-1457. [https://doi.org/10.1016/S0140-6736\(07\)61602-X](https://doi.org/10.1016/S0140-6736(07)61602-X)

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).

Il est strictement interdit de modifier, transformer ou adapter tout ou partie de ce document, ainsi que de l'utiliser à des fins commerciales. 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

Les questions sont les bienvenues.