

Lire et critiquer un article scientifique

Séance 2 sur 3 : Grilles de lecture et résultat principal

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 ← vous êtes ici**
- Séance 3. Les limites méthodologiques

Cours E1-N2-04 Lire et critiquer un article scientifique. Séance 2 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)

2

Grilles de lecture et résultat principal

Séance 2 — Analyser de façon reproductible, interpréter un chiffre



Pourquoi une grille plutôt qu'une "impression" de lecture

Pourquoi

- **Une grille rend l'analyse reproductible** : deux lecteurs qui utilisent la même grille sur le même article aboutissent à des conclusions comparables.
- **Elle protège des raccourcis d'autorité** : le prestige de la revue, la notoriété des auteurs, le nombre de citations ne disent rien de la validité d'une étude particulière.
- **Le facteur d'impact de la revue (IF) est une moyenne de citations à l'échelle d'une revue**. Il ne se transfère pas à l'article lui-même.
- **Elle vous force à examiner ce que vous auriez "sauté"** : le calcul d'effectif, le mode de randomisation, la gestion des données manquantes.
- **Elle produit une trace écrite** : c'est ce qui rend votre analyse discutable en équipe/ en groupe.
- Une grille n'est pas un score. Additionner des items pour produire une note globale est déconseillé par la plupart des concepteurs d'outils.

Sources : Arrêté du 20 février 2026 ; Sterne et al., 2019 ; Shea et al., 2017

Trois familles d'outils que l'on confond constamment

Distinction majeure

Les trois répondent à des questions différentes. Utiliser l'une à la place de l'autre est une erreur de méthode fréquente.

- **1. Standards de rapport (guidelines) :** CONSORT, STROBE, PRISMA, COREQ, TIDieR. **Question posée** : l'article dit-il tout ce qu'il devrait dire ? Ils évaluent la transparence, pas la validité.
- **2. Grilles d'analyse critique :** CASP, JBI, AMSTAR 2 pour les revues systématiques, guide ANAES 2000, qui porte à la fois une grille de lecture et une échelle de niveaux de preuve. **Question posée** : puis-je faire confiance à ce résultat, et est-il applicable chez moi ? Usage pédagogique et professionnel ; questions ouvertes ou cote de confiance globale.
- **3. Outils de risque de biais :** RoB 2, ROBINS-I, QUADAS-2. **Question posée** : quels mécanismes précis ont pu fausser ce résultat, et dans quel sens ? Usage en revue systématique, jugement par domaine.

Point de vigilance : un article peut satisfaire tous les items CONSORT et présenter un risque de biais élevé selon RoB 2 : bien décrire un défaut ne le fait pas disparaître.

Sources : EQUATOR Network, s. d. ; Critical Appraisal Skills Programme, s. d. ; Joanna Briggs Institute, s. d. ; Shea et al., 2017 ; Agence nationale d'accréditation et d'évaluation en santé, 2000 ; Sterne et al., 2019

Les trois questions de toute lecture critique

Structure commune

Quelle que soit la grille utilisée, elle décline ces trois questions.

- **1. Validité interne** : les résultats de cette étude sont-ils crédibles ? Le protocole permet-il d'écarter les explications alternatives (hasard, biais, facteurs de confusion) ?
- **2. Nature et taille de l'effet** : quels sont les résultats ? Quelle est l'ampleur de l'effet, avec quelle précision, sur quel critère de jugement ?
- **3. Validité externe** : ces résultats sont-ils transposables ? À ma population, dans mon contexte de soin, avec mes moyens ? Les bénéfices l'emportent-ils sur les inconvénients et les coûts ?

Point de vigilance : la question 3 est celle que l'on oublie le plus. Un résultat parfaitement valide obtenu en réanimation aux Pays-Bas peut être inutilisable dans un EHPAD français.

Sources : Critical Appraisal Skills Programme, s. d.

CASP et JBI : les deux familles de grilles à connaître

Grilles ouvertes

CASP — Critical Appraisal Skills Programme

- Onze grilles gratuites sur casp-uk.net, sous licence Creative Commons (attribution, usage non commercial, partage à l'identique) pour VOUS AIDER.
- **Une grille par type d'étude** : essai randomisé, revue systématique (dont deux grilles pour les méta-analyses), cohorte, cas-témoins, transversale, qualitatif, précision diagnostique, évaluation économique, règle de prédiction clinique.
- Format : une dizaine de questions, avec des invites de réflexion. Conçu pour l'apprentissage.

JBI — Joanna Briggs Institute

- Outils disponibles sur jbi.global, couvrant un éventail plus large de type d'études.
- Inclut des outils que CASP ne propose pas : études de prévalence, quasi-expérimental, séries de cas, rapports de cas, preuves textuelles (opinion d'expert, texte de politique).
- Format plus fermé : oui / non / peu clair / non applicable.
- Institut de référence internationale en sciences infirmières, adossé à l'Université d'Adélaïde.

Point de vigilance : les deux sont gratuites et en anglais. Utilisez une grille adaptée au design : appliquer la grille « essai randomisé » à une étude qualitative n'a aucun sens.

Sources : *Critical Appraisal Skills Programme, s. d. ; Joanna Briggs Institute, s. d.*

Les quatre outils de référence pour évaluer le risque de biais

Risque de biais et appréciation critique

- **RoB 2 — essais randomisés** : cinq domaines : processus de randomisation ; écarts aux interventions prévues ; données manquantes sur le critère de jugement ; mesure du critère de jugement ; sélection du résultat rapporté. Outil recommandé par le Cochrane Handbook.
- **ROBINS-I — études non randomisées d'interventions** : sept domaines : confusion ; sélection des participants ; classification des interventions ; écarts aux interventions prévues ; données manquantes ; mesure des critères de jugement ; sélection du résultat rapporté.
- **AMSTAR 2 — revues systématiques (outil d'appréciation critique, non de risque de biais)** : seize items, dont sept jugés critiques : protocole a priori, exhaustivité de la recherche, justification des exclusions, évaluation du risque de biais des études incluses, méthodes de méta-analyse, prise en compte du biais dans l'interprétation, évaluation du biais de publication. Il conclut par une cote de confiance globale (élevée, modérée, faible, très faible), sans score.
- **QUADAS-2 — précision diagnostique** : quatre domaines : sélection des patients ; test évalué ; test de référence ; déroulement et calendrier.

Point de vigilance : RoB 2, ROBINS-I et QUADAS-2 ne produisent pas une note globale mais un jugement par domaine, avec sa justification écrite ; AMSTAR 2 conclut par une cote de confiance, sans score. C'est la justification qui a de la valeur, pas l'étiquette finale.

Sources : Sterne et al., 2019 ; Higgins et al., 2024 ; Sterne et al., 2016 ; Shea et al., 2017 ; Whiting et al., 2011

Grille de lecture en dix questions (1/2)

OUTIL DE TRAVAIL

Grille de synthèse construite pour cette UE à partir de principes issus de CASP, des outils JBI, de RoB 2 et de TIDieR. À adapter au design étudié.

#	Question	Où chercher la réponse
1	La question de recherche est-elle clairement formulée et peut-on en identifier les éléments PICO ou PCC, lorsque ce cadre est adapté au design étudié ?	Fin de l'introduction
2	Le design est-il adapté à cette question ?	Méthodes ; design
3	Qui a été inclus, comment, et combien ? La taille d'échantillon a-t-elle été calculée a priori ?	Méthodes ; participants ; diagramme de flux
4	L'intervention et le comparateur sont-ils décrits assez précisément pour être reproduits ?	Méthodes ; intervention (repère : TIDieR)
5	Le critère de jugement principal est-il clairement défini, pertinent et mesuré de façon appropriée ? (outil validé/standardisé ?)	Méthodes ; critères de jugement

Sources : Critical Appraisal Skills Programme (CASP), 2024 . ; Joanna Briggs Institute, s. d. ; Sterne et al., 2019 ; Hoffmann et al., 2014

Grille de lecture en dix questions (2/2)

OUTIL DE TRAVAIL

#	Question	Où chercher la réponse
6	Les groupes étaient-ils comparables au départ ? Sinon, comment cela a-t-il été traité ?	Tableau 1 ; Méthodes ; analyse
7	Combien de participants sont sortis de l'étude, pourquoi, et comment ont-ils été analysés ?	Diagramme de flux ; Méthodes ; population d'analyse
8	Quel est le résultat sur le critère principal, avec sa mesure d'effet et son intervalle de confiance ?	Résultats ; tableau principal
9	Les auteurs déclarent-ils leurs financements et leurs conflits d'intérêts ? L'étude était-elle enregistrée (sur le site clinical trial) ?	Déclarations ; fin du résumé
10	Ce résultat est-il transposable à ma population et à mon contexte de soin ?	Méthodes ; cadre ; Discussion ; limites

Sources : Critical Appraisal Skills Programme, s. d. ; Joanna Briggs Institute, s. d. ; Sterne et al., 2019

3

Identifier et interpréter le résultat principal



Quatre étapes pour lire un résultat

Méthode

- **1. Retrouver le critère** : quel est le critère de jugement principal annoncé dans les Méthodes ? Vérifier qu'il correspond à celui mis en avant dans les Résultats et la Conclusion.
- **2. Identifier la nature** : variable binaire (événement survenu ou non), continue (score, mesure), ou temps jusqu'à événement ? Cette nature oriente le choix de la mesure d'effet .
- **3. Lire la mesure d'effet** : point d'estimation + intervalle de confiance à 95 %. Repérer le seuil de non-effet : 1 pour les rapports, 0 pour les différences.
- **4. Juger la pertinence** : la taille de l'effet est-elle cliniquement importante ? Un résultat statistiquement significatif peut être cliniquement négligeable.

Point de vigilance : les étapes 1 et 4 sont celles que l'on "saute" régulièrement.

Elles sont pourtant celles qui font la différence entre lire un chiffre et comprendre un résultat.

Sources : Hopewell et al., 2025

Quelle mesure pour quel type de variable (1/2)

MESURES D'EFFET

Nature du critère	Mesure d'effet	Lecture	Absence d'effet
Binaire (événement oui/non)	Risque relatif (RR)	Compare le risque entre deux groupes. RR = 0,80 : le risque est diminué de 20 % en relatif.	1
Binaire	Rapport de cotes (OR)	Compare les odds entre deux groupes. OR < 1 : événement moins fréquent dans le groupe étudié ; OR > 1 : plus fréquent. L'OR est proche du RR lorsque l'événement est rare.	1
Binaire	Différence de risque (RD)	Différence absolue entre les deux taux. Elle permet d'apprécier directement l'importance clinique de l'effet.	0
Binaire	Nombre de sujets à traiter (NNT)	Nombre de personnes à prendre en charge pendant une durée donnée pour éviter un événement supplémentaire. Il est calculé à partir de la différence de risque.	Pas de NNT si RD = 0

À retenir : RR et OR décrivent surtout un effet relatif ; la différence de risque et le NNT aident davantage à comprendre ce que cela change concrètement pour les patients.

Sources : Higgins et al., 2024

© 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.

Quelle mesure pour quel type de variable (2/2)

MESURES D'EFFET

Nature du critère	Mesure d'effet	Lecture	Absence d'effet
Continue (score, mesure)	Différence de moyennes (MD)	Différence moyenne entre les groupes, exprimée dans l'unité du critère. Exemple : -0,92 point de pourcentage d'HbA1c.	0
Continue, mais mesurée avec des échelles différentes	Différence de moyennes standardisée (SMD)	Permet de comparer des études utilisant des échelles différentes pour mesurer le même phénomène. Sans unité ; interprétation moins intuitive.	0
Temps jusqu' à un événement	Hazard ratio (HR)	Compare, au cours du suivi, le taux instantané de survenue de l'événement entre les groupes.	1
Événements rapportés à une durée d'exposition	Rate ratio	Compare les taux d'événements entre les groupes. Utile lorsque l'événement peut survenir plusieurs fois, par exemple les chutes.	1

Point de vigilance : le risque relatif impressionne, la différence de risque informe. « Réduction de 50 % » d'un risque qui passe de 2 pour 10 000 à 1 pour 10 000 concerne un patient sur dix mille.

Sources : Higgins et al., 2024 ; Sun et al., 2025

© 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.

L'intervalle de confiance : effet et précision

Intervalle de confiance

Un intervalle de confiance à 95 % apporte trois informations : l'estimation de l'effet, la précision de cette estimation et les valeurs qui restent compatibles avec les données observées.

- **Lecture pratique** : pour un RR, OR ou HR, la valeur de non-effet est 1 ; pour une différence, elle est 0. Si l'IC 95% contient cette valeur, les données restent compatibles avec l'absence de différence.
- Précision : un IC étroit → estimation plus précise ; un IC large → estimation plus incertaine.
- La largeur de l'intervalle est aussi informative que sa position : un intervalle très large signale une étude peu puissante, même quand il exclut le seuil de non-effet.
- **Exemple** : OR = 0,45 [0,22-0,91]. Cela veut dire que l'estimation est en faveur de l'intervention et l'IC n'inclut PAS 1. Toutefois, l'effet réel compatible avec les données peut être relativement modeste ou beaucoup plus important.

À retenir : regarder toujours l'estimation ET son intervalle de confiance. Un chiffre seul ne suffit pas.

Ce que la valeur p dit — et ce qu'elle ne dit pas

Valeur p

Ce qu'elle dit

- La probabilité d'observer un résultat au moins aussi extrême que celui obtenu, si l'hypothèse nulle était vraie et si le modèle statistique était correct.
- Une mesure de compatibilité entre les données et une hypothèse d'absence d'effet.
- Un signal, à interpréter avec le contexte, la taille de l'effet et la qualité du protocole.

Ce qu'elle ne dit pas

- Elle ne donne pas la probabilité que l'hypothèse nulle soit vraie.
- Elle ne mesure pas la taille de l'effet : $p < 0,001$ ne veut pas dire « effet important ».
- Elle ne mesure pas l'importance clinique.
- $p > 0,05$ ne prouve pas l'absence d'effet : cela peut simplement signaler un manque de puissance.
- Le seuil de 0,05 est une convention, pas une frontière naturelle entre le vrai et le faux.

Point de vigilance : l'American Statistical Association a publié en 2016 une déclaration formelle sur les usages abusifs de la valeur p ; Amrhein et al. (2019) ont appelé à renoncer au seuil de significativité comme couperet.

Sources : Wasserstein & Lazar, 2016 ; Amrhein et al., 2019

© 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.

Quatre chiffres réels, quatre lectures

Application

1,04

Rate ratio des chutes

1,068

OR de mortalité

-0,92 %

Différence moyenne d'HbA1c (sous-groupe 4-6 mois)

87,8 %

Hétérogénéité globale I^2

1,04 – Rate ratio des chutes

Barker et al. (2016) : 1,04 [0,78-1,37]. L'IC contient 1. Cela signifie qu' **aucune différence statistiquement significative n'est mise en évidence**. L'intervalle reste toutefois compatible avec une diminution comme avec une augmentation du taux de chutes.

1,068 – OR de mortalité

Aiken et al. (2014) : 1,068 [1,031-1,106]. L'association est statistiquement significative mais de faible amplitude. **Comme il s'agit d'une étude observationnelle, association ne signifie pas causalité.**

-0,92-HbA1c

Sun et al. (2025) : -0,92 point de pourcentage [-1,44 ; -0,41] au suivi intermédiaire. Le résultat est en faveur de l'intervention, mais l'effet varie selon la durée de suivi.

87,8%- I^2

$I^2=87,8\%$: **hétérogénéité importante. Les résultats des études diffèrent fortement entre eux : l'estimation globale doit donc être interprétée avec prudence et il faut rechercher pourquoi les études donnent des résultats différents.**

Sources : Barker et al., 2016 ; Aiken et al., 2014 ; Sun et al., 2025

© 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.

Lire un forest plot en cinq points

Forest plot

Le forest plot représente les résultats des différentes études d'une méta-analyse et, lorsqu'un regroupement est pertinent, leur résultat combiné.

- **Chaque ligne** représente une étude incluse. Le **carré** indique l'estimation de l'effet ; plus il est grand, plus l'étude pèse dans le calcul.
- **Le trait horizontal** représente l'IC 95 % de cette étude. Plus il est long, moins l'estimation est précise.
- **La ligne verticale** représente l'absence d'effet : **1 pour un rapport (RR, OR, HR), 0 pour une différence**. Si l'IC d'une étude traverse cette ligne, cette étude ne met pas en évidence de différence statistiquement significative au seuil correspondant.
- **Le losange** représente l'effet combiné de la méta-analyse ; sa largeur correspond à son IC.
- **L'hétérogénéité**, notamment décrite par I^2 , indique dans quelle mesure les résultats diffèrent entre les études.

Point de vigilance : avant de regarder uniquement le losange, regardez si les études racontent globalement la même histoire. Une forte hétérogénéité impose de comprendre les différences entre les études avant d'appliquer le résultat.

Le graphique en entonnoir

Biais de publication

Le principe

- Chaque étude est représentée par un point : son **effet estimé** est placé sur l'axe horizontal et sa **précision** sur l'axe vertical.
- Lorsque les petites et grandes études donnent des résultats cohérents, le nuage tend à avoir une forme approximativement symétrique.
- Une asymétrie peut faire suspecter que certaines petites études sont manquantes ou donnent des résultats différents.

Les limites de l'interprétation

- **Une asymétrie ne prouve pas un biais de publication.** Elle peut aussi être liée à l'hétérogénéité, à des différences méthodologiques entre petites et grandes études ou simplement au hasard.
- Avec **moins d'une dizaine d'études**, le funnel plot et les tests d'asymétrie sont généralement peu fiables.
- Le **test d'Egger** peut rechercher statistiquement une asymétrie, mais **il ne démontre pas à lui seul un biais de publication.**

Point de vigilance : Le funnel plot est un signal d'alerte : il invite à chercher une explication, il ne permet pas à lui seul de conclure.

Statistiquement significatif $\neq$ important pour le patient

Pertinence

- La significativité statistique répond à une question : les données sont-elles suffisamment incompatibles avec l'hypothèse d'absence d'effet ?
- La pertinence clinique répond à une autre question : l'effet observé est-il suffisamment important pour avoir un intérêt pour le patient ou pour les soins ?
- La **différence minimale cliniquement importante (DMCI)** est la plus petite différence considérée comme suffisamment importante pour avoir un sens clinique.
- Dans une très grande étude, une différence minime peut devenir statistiquement significative **sans être utile en pratique**.
- Inversement, dans une petite étude, un effet potentiellement important peut rester **très incertain**.
- Il faut donc regarder ensemble **la taille de l'effet, son IC et sa pertinence clinique**.

Question du lecteur : cette différence changerait-elle quelque chose pour la personne que je soigne ?

Du risque relatif au nombre de sujets à traiter

CALCUL GUIDÉ

Exemple pédagogique construit pour l'exercice : chiffres fictifs, méthode réelle. Population fictive : 400 résidents à haut risque d'escarre, suivis 6 mois.

Étape	Calcul	Résultat
Risque dans le groupe témoin	24 escarres sur 200 patients	12,0 %
Risque dans le groupe intervention	12 escarres sur 200 patients	6,0 %
Risque relatif (RR)	$6,0 \% \div 12,0 \%$	0,50
Réduction relative du risque	$1 - 0,50$	50 %
Différence de risque (RD)	$12,0 \% - 6,0 \%$	6,0 points
Nombre de sujets à traiter (NNT)	$1 \div 0,06$	≈ 17 patients

Point de vigilance : le même résultat s'énonce « réduction de 50 % du risque » ou « il faut prendre en charge 17 patients pour éviter une escarre ». La seconde formulation est celle qui permet de décider.

Exercice d'application — séance 2

Travail sur article

Le support

- Au choix, deux articles en accès libre :
- Barker et al. (2016) : essai randomisé en grappes (programme 6-PACK de prévention des chutes).
- Sun et al. (2025) : méta-analyse avec forest plots (éducation à l'autogestion du diabète de type 2 menée par des infirmières).

Le travail demandé

- 1. Appliquer la grille en dix questions vue en cours, en répondant par une phrase justifiée et non par oui/non.
- 2. Identifier le résultat sur le critère principal : mesure d'effet, intervalle de confiance, interprétation en une phrase.
- 3. Pour la méta-analyse : décrire le forest plot et commenter l'hétérogénéité.
- 4. Conclure : que permet et que ne permet pas cette étude ?

Point de vigilance : le critère d'évaluation n'est pas le nombre de défauts relevés, mais la justesse de la justification de chacun.

Corrigé type de l'exercice 2

CE QUE L'ON ATTEND (BARKER ET AL., 2016 ; SUN ET AL., 2025)

Grille en dix questions, sur Barker et al. (2016)

- **Q1, PICO** : patients hospitalisés en soins aigus / programme 6-PACK / soins habituels / chutes et chutes avec blessure. **Q2** : la randomisation en grappes convient à une intervention organisationnelle délivrée par service. **Q3** : 24 services, 46 245 admissions, calcul d'effectif présenté. **Q4** : six composantes décrites. **Q5** : deux critères principaux, désignés dans les méthodes. **Q8** : rate ratio 1,04 (IC 95 % 0,78-1,37), l'intervalle contient 1. **Q9** : financements, liens d'intérêt et enregistrement déclarés. **Q10** : hôpitaux australiens ; la transposition se discute service par service.
- **Le résultat sur le critère principal**

Aucun effet démontré. L'intervalle, de 0,78 à 1,37, reste compatible avec une réduction de 22 % comme avec une augmentation de 37 % des chutes : l'essai n'a pas montré d'effet, il n'a pas prouvé l'absence d'effet.

Forest plot et hétérogénéité, sur Sun et al. (2025)

- **Huit essais randomisés.** Pour l'HbA1c, les carrés ne pointent pas tous du même côté et $I^2 = 87,8\%$: l'estimation groupée ne peut pas être retenue telle quelle. L'analyse par délai de suivi (−0,22 à 0-3 mois ; −0,92 à 4-6 mois ; −0,54 au-delà de 6 mois) est une piste d'explication, à traiter comme une hypothèse, non comme un résultat.
- **Conclure**

Barker : un programme correctement mis en œuvre, sans effet mesurable sur les chutes dans ce contexte.

Sun : un effet probable sur l'HbA1c, dont l'ampleur dépend du délai et reste incertaine du fait de l'hétérogénéité.

Sources : Barker et al., 2016 ; Sun et al., 2025 ; Higgins & Thompson, 2002

Vérifions

Auto-évaluation 2

La question

Un essai rapporte : « réduction du risque d'escarre de 40 % dans le groupe intervention (RR = 0,60 ; IC 95 % 0,32-1,12 ; p = 0,11) ». Quelle lecture est correcte ?

- A. L'intervention réduit le risque d'escarre de 40 %, le résultat est démontré
- B. L'intervention est inefficace, puisque $p > 0,05$
- C. Le résultat ne démontre pas d'effet, mais l'intervalle de confiance large reste compatible avec une réduction importante : l'étude manque probablement de puissance
- D. Il faut convertir le RR en OR avant de pouvoir conclure

La réponse

C. L'intervalle de confiance contient 1 : le seuil conventionnel de significativité n'est pas atteint. Mais sa largeur, de 0,32 à 1,12, signifie que les données sont compatibles avec une réduction de 68 % du risque comme avec une augmentation de 12 %. **Conclure à l'inefficacité serait une erreur** : absence de preuve n'est pas preuve d'absence. **La lecture juste consiste à dire que l'essai n'a pas tranché, probablement faute d'effectif suffisant, et à vérifier le calcul d'effectif dans les méthodes.**

Séance 2 — À retenir

Synthèse

- **Trois familles d'outils, trois questions différentes** : rapporter, apprécier, évaluer le risque de biais. Ne les confondez pas.
- **Toute grille décline trois questions** : les résultats sont-ils valides, quels sont-ils, sont-ils transposables ?
- **Le résultat principal se lit en quatre temps** : retrouver le critère, identifier sa nature, lire la mesure d'effet et son intervalle, juger la pertinence clinique.
- **L'intervalle de confiance est plus informatif que la valeur p** : sa position et sa largeur comptent autant l'une que l'autre.
- **Le risque relatif impressionne**, la différence de risque informe. Cherchez toujours l'effet absolu.
- **Significatif n'est pas important, et non significatif n'est pas inefficace.**

Séance 3. Les limites méthodologiques

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)

Agence nationale d'accréditation et d'évaluation en santé. (2000). *Guide d'analyse de la littérature et gradation des recommandations*. https://www.has-sante.fr/jcms/c_434715/

Aiken, L. H., Sloane, D. M., Bruyneel, L., Van den Heede, K., Griffiths, P., Busse, R., Diomidous, M., Kinnunen, J., Kózka, M., Lesaffre, E., McHugh, M. D., Moreno-Casbas, M. T., Rafferty, A. M., Schwendimann, R., Scott, P. A., Tishelman, C., van Achterberg, T., & Sermeus, W. (2014). Nurse staffing and education and hospital mortality in nine European countries: A retrospective observational study. *The Lancet*, 383(9931), 1824-1830. [https://doi.org/10.1016/S0140-6736\(13\)62631-8](https://doi.org/10.1016/S0140-6736(13)62631-8)

Amrhein, V., Greenland, S., & McShane, B. (2019). Scientists rise up against statistical significance. *Nature*, 567(7748), 305-307. <https://doi.org/10.1038/d41586-019-00857-9>

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>

Critical Appraisal Skills Programme. (2024). *CASP checklist: for Randomised Controlled Trials (RCTs)*.

Références bibliographiques (2 sur 4)

Egger, M., Davey Smith, G., Schneider, M., & Minder, C. (1997). Bias in meta-analysis detected by a simple, graphical test. *BMJ*, 315(7109), 629-634. <https://doi.org/10.1136/bmj.315.7109.629>

EQUATOR Network. (s. d.). *Enhancing the QUALity and Transparency Of health Research*. <https://www.equator-network.org/>

Higgins, J. P. T., Thomas, J., Chandler, J., Cumpston, M., Li, T., Page, M. J., & Welch, V. A. (dir.). (2024). *Cochrane handbook for systematic reviews of interventions* (version 6.5). Cochrane. <https://training.cochrane.org/handbook/current>

Higgins, J. P. T., & Thompson, S. G. (2002). Quantifying heterogeneity in a meta-analysis. *Statistics in Medicine*, 21(11), 1539-1558. <https://doi.org/10.1002/sim.1186>

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>

Références bibliographiques (3 sur 4)

Joanna Briggs Institute. (s. d.). *JBI critical appraisal tools*. <https://jbi.global/critical-appraisal-tools>

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>

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>

Sun, J., Fan, Z., Kou, M., Wang, X., Yue, Z., & Zhang, M. (2025). Impact of nurse-led self-management education on type 2 diabetes: A meta-analysis. *Frontiers in Public Health*, 13, 1622988. <https://doi.org/10.3389/fpubh.2025.1622988>

Wasserstein, R. L., & Lazar, N. A. (2016). The ASA statement on p-values: Context, process, and purpose. *The American Statistician*, 70(2), 129-133. <https://doi.org/10.1080/00031305.2016.1154108>

Références bibliographiques (4 sur 4)

Whiting, P. F., Rutjes, A. W. S., Westwood, M. E., Mallett, S., Deeks, J. J., Reitsma, J. B., Leeflang, M. M. G., Sterne, J. A. C., Bossuyt, P. M. M., & QUADAS-2 Group. (2011). QUADAS-2: A revised tool for the quality assessment of diagnostic accuracy studies. *Annals of Internal Medicine*, 155(8), 529-536.
<https://doi.org/10.7326/0003-4819-155-8-201110180-00009>



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.