

Poly de révision EPI 2026

Rosenström et al., Lancet Psychiatry 2025 : questions, corrections détaillées et raisonnement d'examen

Table des matières

Poly de révision EPI 2026	2
Rosenström et al., Lancet Psychiatry 2025 : article, questions, méthodes et réponses détaillées	2
Comment utiliser ce poly	2
1. Carte rapide des grands thèmes 2026	2
2. Briques méthodologiques spécifiques à l'article	3
2.1. Design : effectiveness vs efficacy	3
2.2. Différence brute vs ATE	3
2.3. Ce qu'est TMLE	3
2.4. Le Super Learner	4
2.5. Hypothèses causales : consistance, échangeabilité, positivité	4
2.6. Données manquantes et dropout	5
2.7. Non-pré-enregistrement	5
2.8. Précision, IC95% et puissance a posteriori	5
3. 2026 EPI résolu question par question	6
Article et contexte	6
Extrait de l'article à analyser	7
Ce que l'examineur veut vérifier	7
Thème 1. Design de l'étude	7
Thème 2. Méthodes statistiques	10
Thème 3. Inférence causale et DAG	12
Thème 4. Résultats et taille d'effet	14
Thème 5. Biais et facteurs de confusion	17
Thème 6. Mesures et variables	19
Thème 7. Éthique et réglementaire	21
Thème 8. Santé publique et clinique	22
Thème 9. Critique globale	23
4. Réponses prêtes à l'emploi	24
4.1. Quand il faut discuter la précision plutôt que la puissance	24
4.2. Quand il faut discuter une grande cohorte sans se faire piéger	24
4.3. Quand il faut parler du TMLE sans réciter	24
5. Checklist finale	24
6. Conclusion	24

Poly de révision EPI 2026

Rosenström et al., Lancet Psychiatry 2025 : article, questions, méthodes et réponses détaillées

Document de travail orienté examen : les **35 questions 2026** reprises dans un poly autonome, avec explications pédagogiques, exemples concrets tirés de l'article et réponses prêtes à réutiliser.

Ce poly a un seul objectif : **permettre de répondre entièrement au sujet EPI 2026 à partir de l'article de Rosenström et al.**

La logique est simple :

1. comprendre ce que le papier fait réellement ;
2. identifier ce que chaque question veut tester ;
3. relier la réponse à des passages et chiffres concrets de l'article ;
4. produire une réponse d'examen claire, prudente et défendable.

EPI 2026 TMLE ATE causalité HTML + PDF

! Point central

Le piège principal de l'article 2026 est de le lire comme une "preuve causale forte" simplement parce que la méthode statistique est sophistiquée.

Le bon raisonnement est plus exigeant :

- **méthode moderne** ne veut pas dire **absence de biais** ;
- **grand effectif** ne veut pas dire **comparabilité parfaite** ;
- **IC95% excluant 0** ne veut pas dire **effet cliniquement majeur**.

Comment utiliser ce poly

1. Lire d'abord les briques méthodologiques de la section 2.
2. Revenir ensuite à la section 2026 EPI et travailler les questions dans l'ordre.
3. Pour chaque question, repérer :
 - ce qui vient de l'article ;
 - ce qui relève de l'interprétation méthodologique ;
 - ce qu'il faut écrire dans une copie.
4. Utiliser la section 4 comme banque de formulations prêtes à l'emploi.

1. Carte rapide des grands thèmes 2026

Thème	Ce que l'examineur veut voir
design naturaliste	distinguer effectiveness et efficacy
TMLE / super learner	expliquer sans réciter du jargon vide
causalité	connaître consistence, échangeabilité, positivité
résultats	distinguer différence brute, ATE, précision et pertinence clinique
biais	voir le rôle du dropout, des données manquantes et de la confusion non mesurée

Thème	Ce que l'examineur veut voir
santé publique	savoir ne pas sur-généraliser un résultat finlandais naturaliste

2. Briques méthodologiques spécifiques à l'article

2.1. Design : effectiveness vs efficacy

L'article compare deux modalités réelles de prise en charge :

- iCBT guidée, largement diffusée dans le système public finlandais ;
- fCBT, essentiellement issue de la région Uusimaa.

Ce n'est donc **pas** un essai randomisé.

Le papier documente d'abord l'**effectiveness** en conditions réelles, puis tente d'approcher une question de type efficacy grâce à une méthode causaliste (TMLE).

La phrase importante à retenir est :

Le papier observe des soins réels puis reconstruit une comparaison contrefactuelle ; il n'observe pas directement deux groupes randomisés comparables au départ.

2.2. Différence brute vs ATE

Dans le tableau 2, deux quantités coexistent :

- la **différence observée brute** de changement de PHQ-9 : 1,120
- l'**ATE estimé** par TMLE : 0,745

La différence brute répond à :

qu'a-t-on observé dans les deux groupes tels qu'ils existaient ?

L'ATE répond à :

que se serait-il passé, en moyenne, si ces patients avaient tous reçu iCBT ou tous reçu fCBT ?

L'écart entre 1,120 et 0,745 est déjà un signal pédagogique important : une partie de la différence observée brute vient probablement des **déséquilibres de base** entre groupes.

2.3. Ce qu'est TMLE

Le TMLE (*Targeted Maximum Likelihood Estimator*) combine :

- un modèle d'assignation au traitement ;
- un modèle d'outcome ;
- une étape de "targeting" orientée vers le paramètre causal d'intérêt, ici l'ATE.

Le papier insiste sur deux idées :

- TMLE vise un compromis biais-variance favorable pour le paramètre cible ;
- l'utilisation d'un Super Learner évite de dépendre d'un seul modèle mal spécifié.

Il faut cependant éviter la formule magique :

TMLE n'efface pas les confondeurs non mesurés ; il exploite au mieux les variables observées, sous hypothèses causales fortes.

2.4. Le Super Learner

Le Super Learner est un **empilement d'algorithmes** : régressions généralisées, lasso, elastic net, arbres, boosting, splines, highly adaptive lasso, etc.

Le papier précise que :

- deux super learners sont nécessaires ;
- l'un prédit l'assignation au traitement ;
- l'autre prédit l'outcome ;
- la pondération des apprenants est choisie par validation croisée.

Ce que cela apporte :

- moins d'arbitraire qu'un seul modèle ;
- plus de souplesse face aux relations non linéaires ;
- meilleure robustesse prédictive.

Ce que cela n'apporte pas :

- aucune garantie contre un biais de sélection structurel ;
- aucune mesure directe de variables absentes comme l'affinité numérique ou certaines préférences thérapeutiques.

2.5. Hypothèses causales : consistance, échangeabilité, positivité

Consistance

Elle suppose que "recevoir iCBT" et "recevoir fCBT" ont un sens suffisamment stable pour définir une intervention comparable.

Dans cet article, ce point est délicat parce que les modalités de soins réels diffèrent sur plusieurs dimensions :

- format numérique vs présentiel ;
- organisation de l'offre ;
- intensité de contact ;
- type de thérapeute ;
- contexte régional.

Échangeabilité

Elle suppose qu'après ajustement, il ne reste plus de confondeur non mesuré liant le choix du traitement à l'issue.

C'est la grande hypothèse fragile du papier. Les auteurs ajustent beaucoup, mais ils ne mesurent pas tout.

Positivité

Elle impose qu'à profils comparables, chaque patient ait une probabilité non nulle de recevoir chaque modalité.

Ici, le très fort déséquilibre 5446 vs 388, plus la concentration de fCBT en Uusimaa, rendent cette hypothèse localement plus fragile.

2.6. Données manquantes et dropout

Le papier prend une décision méthodologique forte :

si un patient n'a qu'une seule mesure de PHQ-9, le changement est fixé à 0, ce qui pénalise le dropout.

Cela a un avantage :

- on n'ignore pas les abandons ;
- on évite de ne garder que les completers les plus favorables.

Mais cela a aussi une limite :

- un abandon ne signifie pas toujours absence d'amélioration ;
- il peut refléter aggravation, démotivation, amélioration rapide, ou sortie logistique.

Les auteurs complètent cela par des imputations et plusieurs analyses de sensibilité. La bonne copie doit donc parler de **stratégie sérieuse mais non définitive**.

2.7. Non-pré-enregistrement

Le papier n'est pas pré-enregistré. Cela oblige à discuter :

- liberté analytique a posteriori ;
- risque de sélection des analyses les plus convaincantes ;
- besoin renforcé de transparence et de cohérence entre analyses.

Le fait que les auteurs publient le code et multiplient les sensibilités est un point favorable, mais ne remplace pas un protocole verrouillé à l'avance.

2.8. Précision, IC95% et puissance a posteriori

Le vrai outil principal ici est la **précision**, pas la puissance a posteriori.

Mais si l'examineur force le raisonnement, on peut reconstruire une puissance grossière à partir de l'IC95%.

Formules à connaître :

$$IC95\% = \hat{\theta} \pm 1,96 \times SE$$

Donc :

$$SE \approx \frac{\text{borne sup} - \text{borne inf}}{3,92}$$

Puis :

$$z \approx \frac{\hat{\theta}}{SE}$$

Et, si l'on force le raisonnement de puissance observée :

$$\text{puissance observée} \approx \Phi(z - 1,96)$$

Application détaillée à l'ATE principal

Données de l'article :

$$ATE = 0,745$$

$$IC95\% = [0,156 ; 1,334]$$

Largeur de l'intervalle :

$$1,334 - 0,156 = 1,178$$

Erreur standard reconstruite :

$$SE \approx \frac{1,178}{3,92} \approx 0,300$$

Score de Wald grossier :

$$z \approx \frac{0,745}{0,300} \approx 2,48$$

Puissance observée grossière :

$$\Phi(2,48 - 1,96) = \Phi(0,52) \approx 0,70$$

Interprétation :

- il existe un signal statistique compatible avec un niveau d'information raisonnable ;
- mais la vraie discussion reste la **taille d'effet modeste** et la **validité causale** ;
- on ne doit donc jamais transformer cette puissance a posteriori en "preuve de qualité" de l'étude.

i Ce qu'il faut écrire à l'examen

Dans cette étude, la puissance a posteriori n'est pas le vrai sujet. Si on la calcule malgré tout à partir de l'IC95%, on trouve un ordre de grandeur d'environ 70%. Mais la discussion prioritaire porte surtout sur la précision de l'estimation, l'ampleur clinique de l'effet, et la plausibilité des hypothèses causales.

3. 2026 EPI résolu question par question**Article et contexte**

Article support : **Rosenström et al., Lancet Psychiatry 2025**, comparaison naturaliste finlandaise entre **internet-delivered CBT** guidée (iCBT) et **face-to-face CBT** (fCBT) pour la dépression.

- cohorte rétrospective sur registres de soins ;
- 5834 patients analysés après exclusion des données basales manquantes ;
- environ 5446 patients iCBT contre 388 patients fCBT ;
- outcome principal : changement de PHQ-9 ;
- estimateur principal : **ATE par TMLE** ;
- résultat central : ATE = 0,745 avec IC95% [0,156 ; 1,334] ;

- point critique : très belle sophistication méthodologique, mais absence de randomisation et forte dissymétrie des groupes.

Extrait de l'article à analyser

Ce que l'examineur veut vérifier

- lecture critique d'un papier observationnel très technique ;
- capacité à expliquer TMLE, ATE, causalité et limites sans réciter du vocabulaire creux ;
- aptitude à relier une question méthodologique à des chiffres concrets de l'article ;
- capacité à discuter à la fois la **signification statistique**, la **pertinence clinique** et la **validité causale**.

Thème 1. Design de l'étude

Question 1

Il s'agit d'une cohorte rétrospective utilisant des registres naturalistes. Quels sont les avantages et les inconvénients de ce type de données par rapport à un essai contrôlé randomisé ?

Ce que l'examineur veut vérifier : savez-vous opposer la validité externe d'un registre naturaliste à la validité interne d'un essai randomisé, sans caricaturer ni l'un ni l'autre ?

Réponse détaillée

L'avantage évident d'un registre naturaliste est l'accès à une **population très large et très proche du soin réel**. Ici, les auteurs peuvent partir de 32 343 entrées et aboutir à plus de 5800 patients analysés, ce qui serait coûteux et lent dans un essai randomisé. On gagne donc en **validité externe** : on voit comment les soins fonctionnent dans la vraie vie, avec des patients adressés dans le système de santé finlandais tel qu'il existe vraiment.

Mais le prix à payer est important : il n'y a **pas de randomisation**. Les patients orientés vers iCBT et fCBT ne sont probablement pas comparables au départ. Le tableau 1 montre déjà des différences majeures : âge moyen plus élevé en fCBT (40,37 ans contre 34,49), plus de visites psychiatriques antérieures (43,89 contre 16,29), PHQ-9 initial plus bas (11,57 contre 13,31), et profil socioprofessionnel différent. Cela suggère que les groupes reflètent des parcours de soins distincts, pas deux bras d'essai échangés aléatoirement.

Par rapport à un essai contrôlé randomisé, la bonne conclusion est donc : **meilleure représentativité clinique, mais moins bonne protection contre la confusion et le biais de sélection**. Une bonne copie ne dit pas seulement "registre = bien" ou "pas randomisé = mauvais". Elle dit : **excellent pour documenter l'effectiveness, insuffisant pour conclure causalement sans hypothèses fortes**.

Réponse type courte

Cette cohorte naturaliste a pour force principale sa validité externe et sa grande taille, mais son principal défaut est l'absence de randomisation. Elle décrit très bien le soin réel, mais elle reste exposée à la confusion résiduelle et à des groupes initialement non comparables.

Question 2

Expliquez le flowchart : comment les 32 343 entrées ont-elles été réduites à 5 834 patients analysés ? Que révèle ce processus sur la représentativité de l'échantillon ?

Ce que l'examineur veut vérifier : savez-vous lire un flowchart comme une pièce méthodologique centrale et non comme une simple formalité graphique ?

Réponse détaillée

Le flowchart montre d'abord que les auteurs n'analysent pas "tous les patients dépressifs traités", mais seulement ceux qui satisfont un enchaînement de conditions techniques et cliniques. Sur 32 343 entrées évaluées, une grande masse est exclue pour des raisons liées au diagnostic, à la disponibilité des mesures PHQ-9, au respect des critères de la comparaison, ou à la structure même des deux registres. Après ce tri, il reste 392 patients fCBT et 5467 patients iCBT pour les analyses naturalistes, puis 25 sont encore exclus des analyses contrefactuelles pour données basales manquantes, ce qui donne 5834 patients finaux.

Ce processus révèle deux choses. Premièrement, l'échantillon final est **cliniquement pertinent**, car il correspond à des patients pour lesquels la comparaison iCBT / fCBT est méthodologiquement possible. Deuxièmement, il n'est pas forcément représentatif de tous les patients initialement vus dans le système : il représente surtout les **patients avec données exploitables** et appartenant aux bons circuits de soins. Cela expose à un biais de sélection subtil : on généralise plus facilement au sous-ensemble des patients réellement captés et mesurés qu'à l'ensemble des patients dépressifs pris en charge.

L'examineur attend ici qu'on dise explicitement qu'un flowchart sert à juger **qui entre réellement dans l'analyse**. La réduction drastique des effectifs n'est pas seulement une perte numérique ; elle dit quelque chose sur la population réellement décrite par l'étude.

Réponse type courte

Le flowchart réduit 32 343 entrées à 5834 patients analysés en éliminant les dossiers non comparables ou insuffisamment documentés. L'échantillon final reste cliniquement utile, mais il représente surtout les patients avec données exploitables, ce qui peut limiter la représentativité globale.

Question 3

Expliquez la distinction entre "efficacy" et "effectiveness" telle qu'elle est utilisée dans cet article. En quoi cette étude apporte-t-elle des preuves sur les deux ?

Réponse détaillée

L'*effectiveness* désigne l'efficacité en conditions réelles : comment une intervention fonctionne dans la vraie vie, avec les contraintes d'organisation, les abandons, les hétérogénéités de patients et les circuits ordinaires de soins. L'*efficacy*, elle, renvoie à une comparaison plus "propre", plus proche de ce qu'on verrait si deux traitements étaient évalués dans des conditions très contrôlées.

Cet article apporte d'abord une information forte sur l'*effectiveness* parce qu'il observe ce qui se passe réellement dans le système finlandais. Les patients iCBT et fCBT ont reçu les soins tels qu'ils existent vraiment. Mais les auteurs ne s'arrêtent pas là : avec TMLE, ils tentent de reconstruire une comparaison contrefactuelle plus proche d'une logique d'*efficacy*. Cela ne les transforme pas en essai randomisé ; cela signifie seulement qu'ils essaient d'aller au-delà de la pure description brute.

La réponse d'examen attendue est donc nuancée : **oui, le papier informe sur l'effectiveness ; il essaie aussi d'approcher une question d'efficacy ; mais il n'atteint pas le niveau de preuve d'un essai contrôlé randomisé**. C'est précisément cette tension qui fait l'intérêt méthodologique du papier.

Réponse type courte

Le papier documente d'abord l'*effectiveness*, car il analyse des soins réels. Grâce au TMLE, il essaie ensuite d'approcher une question d'*efficacy* contrefactuelle, sans toutefois égaler la force causale d'un essai randomisé.

Question 4

L'étude n'a pas été pré-enregistrée. Quelles sont les implications méthodologiques de ce choix et comment les auteurs tentent-ils d'y remédier ?

Réponse détaillée

L'absence de pré-enregistrement ouvre un espace de liberté analytique plus grand : choix des modèles, des covariables, des restrictions, des sensibilités et des présentations des résultats peuvent être faits après avoir vu les données. Cela augmente le risque de HARKing et de sélection a posteriori des analyses les plus convaincantes. Dans un papier aussi riche méthodologiquement, ce point n'est pas anecdotique.

Les auteurs essaient cependant de limiter ce problème de plusieurs façons. Ils publient des analyses de sensibilité nombreuses (G-computation, AIPTW, versions simplifiées du TMLE, restriction Uusimaa, imputations), explicitent leur code et montrent une cohérence générale des résultats. Par exemple, l'ATE principal est 0,745, l'AIPTW-ATE est 0,735, et la version simplifiée du TMLE tourne autour de 0,790. Cette convergence aide à rendre l'ensemble plus crédible.

La bonne réponse n'est donc ni "pas pré-enregistré = invalide", ni "beaucoup d'analyses = tout va bien". Il faut dire : **c'est une faiblesse réelle, partiellement compensée par la transparence et la robustesse de l'ensemble des analyses.**

Réponse type courte

Le non-pré-enregistrement fragilise la crédibilité confirmatoire de l'étude en laissant plus de place aux choix analytiques a posteriori. Les auteurs réduisent partiellement cette faiblesse par la transparence et la cohérence de nombreuses analyses de sensibilité, sans l'annuler complètement.

Question 5

Quelles sont les différences entre le groupe iCBT et le groupe fCBT au niveau du traitement lui-même ? En quoi ces différences compliquent-elles l'interprétation des résultats ?

Réponse détaillée

Le papier ne compare pas un facteur simple du type "même thérapie, seul le support change". Il compare deux modalités de prise en charge réelles, qui diffèrent simultanément sur plusieurs plans : le format (internet vs présentiel), le territoire de soins (Uusimaa pour fCBT), l'organisation du parcours, la standardisation des séances, l'accessibilité, et probablement la façon dont les patients sont orientés vers l'une ou l'autre modalité.

Concrètement, cela signifie que l'effet attribué au traitement peut en réalité refléter un **paquet causal** plus large. Par exemple, un patient jeune, plus autonome numériquement et moins lourd sur le plan psychiatrique peut être plus facilement orienté vers iCBT. Or ce profil peut influencer la réponse indépendamment du mode thérapeutique. Le papier essaie d'ajuster ces différences, mais il ne peut pas garantir que tout a été mesuré.

La bonne interprétation est donc de dire qu'on compare des **trajectoires de soins réelles** et non deux interventions parfaitement isolées. Cette phrase est essentielle à l'examen, car elle montre qu'on a compris pourquoi l'inférence causale reste imparfaite malgré la sophistication statistique.

Réponse type courte

Les groupes diffèrent non seulement par le support de thérapie, mais aussi par le contexte de soins, l'orientation des patients et probablement leur profil clinique. Cela complique l'interprétation car l'effet observé peut refléter un ensemble de différences de prise en charge, pas uniquement le format iCBT versus fCBT.

Thème 2. Méthodes statistiques

Question 6

Expliquez ce qu'est le TMLE et pourquoi il est supérieur à une régression multivariée classique ou à un score de propension simple pour estimer l'ATE.

Réponse détaillée

Le TMLE est un estimateur causal ciblé vers un paramètre précis, ici l'ATE. Il ne se contente pas d'ajuster "comme on peut" une régression ; il combine un modèle de traitement et un modèle d'outcome, puis ajuste l'estimation finale pour optimiser le biais-variance du paramètre recherché. C'est pour cela que les auteurs le présentent comme un outil moderne de **counterfactual causal inference**.

Par rapport à une régression multivariée classique, le TMLE a deux avantages pédagogiquement importants. D'abord, il est moins dépendant d'une seule spécification de modèle, surtout quand il est couplé à un Super Learner. Ensuite, il est explicitement construit pour viser l'ATE, pas seulement une association conditionnelle. Par rapport à un score de propension simple, il est également plus riche, car il ne fait pas reposer toute l'inférence sur le seul modèle d'assignation au traitement.

Mais il faut rester précis : "supérieur" ne signifie pas "incontestable". Si des variables décisives ne sont pas mesurées, comme l'affinité numérique ou certains motifs fins d'orientation vers fCBT, le TMLE ne peut pas les inventer. En copie, il faut donc dire : **outil très puissant pour exploiter les données observées, sans supprimer les limites liées aux données non observées**.

Réponse type courte

Le TMLE combine un modèle de traitement et un modèle d'outcome pour estimer directement un paramètre causal comme l'ATE. Il est plus robuste qu'une régression unique ou qu'un score de propension isolé, mais il ne corrige pas les confondeurs non mesurés.

Question 7

Qu'est-ce qu'un Super Learner ? Expliquez le principe de validation croisée utilisé et son intérêt.

Réponse détaillée

Un Super Learner est une méthode d'agrégation d'algorithmes. Au lieu de choisir arbitrairement un seul modèle, on met en concurrence plusieurs candidats : régressions linéaires, lasso, elastic net, arbres, boosting, splines, highly adaptive lasso, etc. La validation croisée sert à estimer leurs performances prédictives puis à construire une combinaison pondérée qui se comporte mieux, en théorie, que beaucoup de choix naïfs isolés.

Dans l'article, cet outil est particulièrement pertinent parce que les relations entre covariables, assignation au traitement et outcome sont probablement non linéaires et complexes. Par exemple, l'effet combiné de l'âge, des antécédents psychiatriques, du statut social et des prescriptions antérieures peut être difficile à modéliser proprement avec une régression simple. Le Super Learner essaie justement de ne pas imposer trop tôt une structure rigide.

L'intérêt méthodologique est réel, mais la bonne copie ajoute toujours une limite : une très bonne prédiction n'est pas une preuve de causalité. Le modèle d'assignation au traitement a ici une AUC d'environ 0,77, ce qui montre qu'il est informatif, mais cela ne dit rien à lui seul sur la disparition de toute confusion résiduelle.

Réponse type courte

Le Super Learner combine plusieurs algorithmes au lieu d'en choisir un seul. La validation croisée sert

à pondérer les meilleurs apprenants et à réduire le risque de mauvaise spécification, sans pour autant éliminer les problèmes de causalité non observée.

Question 8

Dans l'article, les auteurs mentionnent des analyses de sensibilité incluant G-computation, AIPTW, simplification du TMLE et restriction à Uusimaa. Quel est l'intérêt de cette stratégie ?

Réponse détaillée

L'intérêt est de vérifier que le résultat principal n'est pas entièrement fabriqué par un seul choix méthodologique. Si un papier observationnel moderne ne présentait qu'une estimation principale sans aucune sensibilité, sa crédibilité serait faible. Ici, les auteurs montrent au contraire que le message général persiste lorsqu'on change raisonnablement d'angle d'analyse.

Les chiffres sont parlants : l'ATE principal est 0,745, l'AIPTW-ATE est 0,735, la version simplifiée du TMLE est proche de 0,790, et la restriction à Uusimaa donne encore un effet positif (0,636) même si l'intervalle devient compatible avec 0. Cela ne prouve pas la causalité, mais cela suggère que le signal n'est pas uniquement un artefact d'un pipeline unique.

La bonne réponse d'examen est donc : **les sensibilités ne sont pas un supplément décoratif; elles sont une condition de crédibilité dans une étude observationnelle de ce type.**

Réponse type courte

Ces analyses de sensibilité testent la robustesse du résultat à d'autres méthodes et à d'autres sous-populations. Elles renforcent la crédibilité du signal sans transformer l'étude en preuve causale définitive.

Question 9

Les auteurs utilisent la validation croisée imbriquée (nested cross-validation). Pourquoi est-ce important ?

Réponse détaillée

La validation croisée imbriquée sert à éviter une forme subtile d'optimisme : celle qui consiste à choisir un modèle sur des données puis à l'évaluer sur ces mêmes données ou sur un découpage insuffisamment séparé. En séparant mieux la phase de tuning et la phase d'évaluation, on obtient une estimation plus honnête de la performance prédictive.

Dans ce papier, c'est particulièrement important parce que les auteurs utilisent des méthodes de machine learning assez riches. Plus les outils sont flexibles, plus le risque d'overfitting augmente si l'on n'encadre pas bien leur apprentissage. La validation croisée imbriquée agit donc comme un garde-fou méthodologique.

La bonne copie doit toutefois éviter de s'arrêter au mot-clé. Il faut ajouter : **cela améliore la qualité prédictive de la procédure, mais ne remplace pas la discussion causale sur les variables absentes.**

Réponse type courte

La validation croisée imbriquée réduit le risque d'optimisme artificiel lié au choix et à l'évaluation des modèles. Elle renforce la rigueur prédictive du Super Learner, sans résoudre à elle seule les problèmes de causalité.

Question 10

Comment l'imputation multiple (superMICE) a-t-elle été utilisée dans cette étude ? Quels sont ses intérêts et ses limites ?

Réponse détaillée

L'article propose des analyses imputées pour tester des scénarios moins sévères que la règle principale de changement nul en cas d'unique mesure PHQ-9. Les auteurs utilisent superMICE pour générer des valeurs plausibles des mesures manquantes, puis recalculent des ATE dans des scénarios où tous les patients auraient complété, ou au moins disposeraient d'une seconde mesure.

Cette stratégie est utile parce qu'elle permet de voir si le message général dépend entièrement de la manière de gérer les manquants. Les résultats imputés sont d'ailleurs plus élevés (1,277 et 1,070 selon le scénario), ce qui montre que la règle principale est plutôt conservatrice pour iCBT. C'est une information méthodologique intéressante.

Mais là encore, il faut rester prudent : une imputation multiple repose sur des hypothèses implicites sur la structure des données manquantes. Si les abandons sont fortement liés à des facteurs non observés, comme une démotivation particulière ou une mauvaise adéquation avec le format de soin, aucune imputation ne "répare" complètement le problème.

Réponse type courte

L'imputation multiple sert ici à tester la sensibilité du résultat principal à d'autres hypothèses sur les données manquantes. C'est une vraie force, mais elle reste dépendante d'hypothèses sur le mécanisme de manque et ne supprime pas tout biais potentiel lié au dropout.

Thème 3. Inférence causale et DAG

Question 11

Proposez et justifiez un DAG représentant les relations causales plausibles dans l'article.

Réponse détaillée

Le DAG minimal doit placer plusieurs variables en amont du **choix de traitement** et de l'**outcome**. Il faut au moins y faire figurer :

- âge ;
- sexe ;
- sévérité dépressive initiale ;
- anxiété (OASIS) ;
- antécédents psychiatriques ;
- statut social et professionnel ;
- région ;
- éventuellement affinité numérique et préférences thérapeutiques.

Ces variables peuvent influencer à la fois la probabilité d'être orienté vers iCBT ou fCBT et la trajectoire ultérieure de PHQ-9. Le traitement influence ensuite le changement symptomatique, mais aussi possiblement le dropout, qui devient alors une variable post-traitement délicate.

L'exercice du DAG a ici un intérêt pédagogique majeur : il montre visuellement que la difficulté du papier ne tient pas seulement au modèle statistique, mais au grand nombre de **chemins de confusion plausibles**. Une bonne copie ne se contente donc pas de lister des flèches ; elle explique que le TMLE essaie de bloquer ces chemins, sans garantir que tout a été mesuré.

Réponse type courte

Le DAG doit placer les variables cliniques, sociales et régionales comme causes communes du traitement et de l'outcome. Son intérêt principal est de rendre visible la masse de chemins de confusion que le TMLE

tente d'ajuster, sans pouvoir prouver qu'aucun confondateur important n'a été oublié.

Question 12

Expliquez les hypothèses causales nécessaires au TMLE dans ce contexte : consistance, échangeabilité, positivité.

Réponse détaillée

La **consistance** suppose que l'on sait ce que signifie "recevoir iCBT" et "recevoir fCBT" de façon suffisamment stable. Ce point n'est pas trivial ici, car iCBT et fCBT ne sont pas deux interventions strictement identiques hormis le support ; elles correspondent à des offres de soin organisées différemment.

L'**échangeabilité** suppose qu'après ajustement sur les covariables observées, il ne reste plus de facteurs non mesurés qui influencent à la fois le choix du traitement et le résultat. C'est la grande hypothèse non vérifiable du papier. Les auteurs ajustent beaucoup, mais ils ne peuvent pas mesurer parfaitement des dimensions comme l'affinité numérique, la motivation fine, les préférences du thérapeute ou du patient.

La **positivité** impose que pour des profils comparables, les deux traitements restent possibles. Or le déséquilibre 5446 vs 388 et la concentration de fCBT dans Uusiaa suggèrent que certains profils ont beaucoup plus de chances d'être vus dans un bras que dans l'autre. Cela ne détruit pas forcément l'analyse, mais affaiblit localement sa robustesse.

Réponse type courte

La consistance, l'échangeabilité et la positivité sont toutes nécessaires au TMLE. Dans cet article, l'échangeabilité et la positivité sont les hypothèses les plus fragiles, à cause de la forte dissymétrie des groupes et de possibles confondateurs non mesurés.

Question 13

Qu'est-ce qu'une analyse de contrôle négatif ? Comment aurait-elle pu être utile ici ?

Réponse détaillée

Une analyse de contrôle négatif consiste à choisir une exposition ou un outcome qui ne devrait pas être causalement lié à l'effet étudié, mais qui partagerait idéalement les mêmes structures de biais et de confusion. Si une association apparaît quand elle ne devrait pas, cela suggère qu'il subsiste un biais caché.

Dans ce papier, un contrôle négatif aurait pu être utile pour tester si la différence iCBT / fCBT produit artificiellement des associations dans des domaines où elle ne devrait pas avoir d'effet plausible à court terme. Cela aurait été particulièrement intéressant vu la crainte d'une sélection structurelle des patients.

L'absence de contrôle négatif n'invalide pas l'article, mais une bonne réponse signale que cela aurait constitué un garde-fou supplémentaire contre la confusion résiduelle.

Réponse type courte

Un contrôle négatif aurait permis de tester l'existence d'un biais caché en cherchant une association là où aucun effet causal n'est plausible. Dans un papier aussi exposé à la confusion par indication, cela aurait été un renfort méthodologique utile.

Question 14

Qu'entend-on par « triangulation causale » dans cet article et dans la littérature ?

Réponse détaillée

La triangulation causale consiste à confronter plusieurs approches imparfaites, mais biaisées différemment, pour voir si elles pointent vers le même message. L'idée n'est pas qu'une méthode "gagne", mais qu'une convergence de résultats obtenus autrement augmente la plausibilité causale.

Dans ce papier, on peut parler de triangulation entre :

- comparaison naturaliste brute ;
- estimation causale par TMLE ;
- analyses de sensibilité (AIPTW, G-computation, variantes simplifiées) ;
- confrontation implicite à la littérature d'essais antérieurs.

Si toutes ces approches suggèrent que iCBT n'est pas moins bonne, voire légèrement meilleure, le message devient plus crédible qu'une seule estimation isolée. Mais il reste toujours conditionné à la qualité des données et aux hypothèses du cadre causal.

Réponse type courte

La triangulation causale consiste à faire converger plusieurs approches imparfaites plutôt qu'à faire reposer toute la preuve sur un seul modèle. Ici, elle renforce la plausibilité du résultat sans supprimer les limites d'une étude observationnelle.

Question 15

La question contrefactuelle posée dans l'article est : « quel serait l'ATE si on pouvait allouer tous les patients à iCBT ou tous à fCBT ? ». Que faut-il comprendre exactement ?

Réponse détaillée

Il faut comprendre que les auteurs ne s'intéressent pas seulement à la différence moyenne observée entre les deux groupes existants. Ils veulent approcher une question plus ambitieuse : **quel aurait été, en moyenne, le changement de PHQ-9 si l'on avait pu imposer l'une ou l'autre modalité à tous ces patients ?**

Cette reformulation est cruciale, car elle déplace le débat. On ne demande plus simplement "qui allait mieux dans les données observées ?", mais "quel est l'effet moyen d'une allocation hypothétique ?". C'est précisément le cœur de l'inférence contrefactuelle.

En examen, il faut bien montrer qu'on a compris ce saut conceptuel. Sinon, on risque de réduire le papier à une simple comparaison descriptive de deux moyennes, ce qu'il n'est pas.

Réponse type courte

L'ATE ne décrit pas seulement ce qui a été observé ; il cherche à approcher ce qui se serait passé si, en moyenne, les mêmes patients avaient tous reçu l'une puis l'autre modalité. C'est donc une quantité contrefactuelle et pas une simple différence brute.

Thème 4. Résultats et taille d'effet

Question 16

L'ATE estimé est de 0,745 point PHQ-9 (IC95% : 0,156–1,334). Interprétez ce résultat sous l'angle statistique et clinique.

Ce que l'examineur veut vérifier : savez-vous distinguer "statistiquement compatible avec un effet" et "cliniquement grand" ? Et savez-vous reconstruire proprement la précision du résultat ?

Réponse détaillée

Statistiquement, l'intervalle de confiance 0,156 à 1,334 exclut 0, donc le papier observe un signal compatible avec une meilleure réduction des symptômes sous iCBT. Si l'on reconstruit l'erreur standard à partir de l'intervalle :

$$SE \approx \frac{1,334 - 0,156}{3,92} = \frac{1,178}{3,92} \approx 0,300$$

Puis :

$$z \approx \frac{0,745}{0,300} \approx 2,48$$

Ce z confirme que le résultat est compatible avec un effet non nul. Si l'on force une puissance observée grossière :

$$\Phi(2,48 - 1,96) = \Phi(0,52) \approx 0,70$$

On obtient donc un ordre de grandeur d'environ 70%, ce qui est raisonnable sans être impressionnant. Mais surtout, ce calcul ne doit pas faire oublier l'essentiel : **l'effet clinique moyen reste modeste**. Un gain de 0,745 point sur une échelle PHQ-9 n'est pas négligeable, mais ce n'est pas une rupture thérapeutique majeure.

Autrement dit : le résultat est **statistiquement crédible**, mais l'amplitude clinique est **faible à modérée**. La bonne copie doit faire vivre cette double lecture, sans tomber ni dans le triomphalisme ni dans le dénigrement.

Réponse type courte

L'ATE de 0,745 avec IC95% [0,156 ; 1,334] indique un signal statistiquement compatible avec une meilleure amélioration sous iCBT. En revanche, l'effet moyen reste cliniquement modeste. Si l'on force un calcul de puissance observée, on trouve un ordre de grandeur d'environ 70%, mais la vraie discussion reste la précision et la validité causale.

Question 17

Expliquez la différence entre l'ATE estimé (0,745) et la différence observée brute (1,120).

Réponse détaillée

La différence brute 1,120 compare simplement les deux groupes tels qu'ils existent dans les données. Elle mélange donc :

- l'effet du traitement ;
- les différences de structure entre groupes ;
- les biais de sélection ;
- la confusion par indication.

L'ATE de 0,745, lui, essaie de corriger une partie de ces déséquilibres. Le fait qu'il soit nettement plus petit que la différence brute est très instructif. Cela suggère qu'une partie non négligeable de l'écart initial venait du fait que les deux groupes n'étaient pas comparables au départ. On voit d'ailleurs dans le tableau 1 que les patients fCBT sont plus âgés, plus chargés en antécédents psychiatriques et ont un PHQ-9 initial plus bas.

La bonne lecture n'est donc pas "le modèle minimise le résultat", mais plutôt : **la différence brute surestimait probablement l'effet causal réel**.

Réponse type courte

La différence brute 1,120 reflète à la fois traitement et déséquilibres initiaux. L'ATE 0,745 tente de corriger ces déséquilibres; sa valeur plus faible suggère qu'une partie de l'écart brut provenait de la non-comparabilité des groupes.

Question 18

Discutez le problème du dropout dans cette étude. Comment les auteurs le gèrent-ils et dans quelle mesure est-ce convaincant ?

Réponse détaillée

Le dropout est un problème central ici parce que l'outcome repose sur la disponibilité d'au moins une mesure de suivi de PHQ-9. Si un patient disparaît du suivi, cela peut refléter plusieurs réalités très différentes : aggravation, amélioration rapide, démotivation, contrainte logistique, ou inadéquation au format proposé.

Les auteurs choisissent une stratégie principale conservatrice : si une seule mesure PHQ-9 est disponible, le changement est fixé à 0. Cette règle pénalise les abandons et évite de surestimer artificiellement l'efficacité des modalités avec plus de complétants motivés. Ils ajoutent ensuite des analyses imputées, qui donnent des effets plus élevés (1,277 ou 1,070 selon le scénario), montrant que le résultat dépend partiellement de la gestion des manquants.

Cette stratégie est méthodologiquement sérieuse, mais pas totalement décisive. Pourquoi ? Parce qu'elle repose encore sur des hypothèses sur le mécanisme de manque. Si les patients qui abandonnent sont profondément différents pour des raisons non mesurées, l'incertitude persiste. La bonne copie doit donc dire : **gestion rigoureuse et prudente du problème, mais impossibilité de garantir que le biais de dropout est entièrement neutralisé.**

Réponse type courte

Le dropout est critique car il touche directement la mesure du PHQ-9. Les auteurs le gèrent d'abord par une règle conservatrice de changement nul, puis par des imputations de sensibilité. C'est une stratégie solide, mais elle ne supprime pas complètement l'incertitude liée aux données manquantes.

Question 19

Les tailles d'effet intra-groupe sont importantes. Pourquoi ne faut-il pas les confondre avec l'effet comparatif principal ?

Réponse détaillée

Les évolutions intra-groupe peuvent être impressionnantes, mais elles répondent à une autre question. Voir un PHQ-9 moyen baisser dans un groupe ne dit pas automatiquement que ce traitement est supérieur à l'autre. Une amélioration au sein d'un groupe peut refléter :

- l'effet du soin ;
- la régression vers la moyenne ;
- l'évolution naturelle de l'épisode ;
- la sélection des patients qui complètent ;
- l'effet commun aux deux prises en charge.

Dans le papier, les auteurs rapportent par exemple une baisse moyenne de 4,8 points parmi certains complètes iCBT, ce qui peut paraître très fort. Mais ce n'est pas l'effet comparatif causal d'intérêt. Le vrai paramètre principal est l'ATE entre modalités, pas la baisse observée dans un seul groupe.

La bonne réponse doit donc rappeler que **les résultats intra-groupe sont descriptifs, tandis que la question scientifique principale est comparative et contrefactuelle.**

Réponse type courte

Une amélioration importante au sein d'un groupe ne démontre pas sa supériorité comparative. Le paramètre principal ici est l'ATE entre iCBT et fCBT, pas la baisse moyenne observée dans un seul bras.

Question 20

Pourquoi les auteurs ont-ils choisi d'exclure le PHQ-9 baseline de leurs modèles de prédiction principaux ?

Réponse détaillée

Le point technique est subtil. L'outcome principal est le **changement** de PHQ-9, donc inclure systématiquement le score initial dans les modèles risque de créer une dépendance mathématique artificielle entre la valeur de départ et la variation observée, surtout en présence d'erreur de mesure. Les auteurs veulent éviter cette forme de "coupling" entre baseline et outcome dérivé.

Ils montrent néanmoins que ce choix n'est pas neutre. Quand on réintroduit le PHQ-9 de base dans certaines analyses de sensibilité, l'ATE peut fortement s'atténuer, jusqu'à devenir proche de 0 dans certains scénarios. Ce point est pédagogiquement très fort : il montre que même une décision techniquement justifiable peut déplacer matériellement l'effet estimé.

La bonne copie n'affirme donc pas que les auteurs "ont raison" ou "ont tort", mais qu'ils font un choix cohérent avec leur définition de l'outcome, tout en laissant voir que ce choix influence les résultats.

Réponse type courte

Les auteurs excluent le PHQ-9 initial pour éviter une dépendance artificielle entre baseline et changement. Ce choix est méthodologiquement défendable, mais il n'est pas neutre, puisque l'inclusion du score initial atténue fortement l'effet dans certaines analyses de sensibilité.

Thème 5. Biais et facteurs de confusion

Question 21

En épidémiologie les facteurs de confusion non mesurés sont une hantise. Quelle stratégie générale les auteurs utilisent-ils pour limiter ce problème ?

Réponse détaillée

Les auteurs adoptent une stratégie cumulative plutôt qu'un argument unique d'autorité. Ils utilisent :

- de nombreuses covariables issues de plusieurs registres ;
- un TMLE avec Super Learner ;
- une modélisation séparée du traitement et de l'outcome ;
- des analyses de sensibilité de plusieurs types ;
- une restriction régionale (Uusiaa) ;
- des imputations pour les données manquantes.

Cette accumulation d'outils réduit probablement une grande partie de la confusion **observée**. Mais la question de l'examineur porte sur la confusion **non mesurée**. Or aucune sophistication ne permet d'ajuster ce qui n'a pas été enregistré : affinité numérique, motivation, préférences thérapeutiques, appréciation clinique fine du thérapeute, etc.

La bonne réponse doit donc dire : **oui, stratégie sérieuse et moderne ; non, pas de garantie d'absence de confusion résiduelle.**

Réponse type courte

Les auteurs réduisent la confusion par une stratégie large : nombreuses covariables, TMLE, super learner, sensibilités, restrictions régionales et imputations. Cela diminue la confusion observée, mais n'élimine pas les confondeurs non mesurés.

Question 22

Les patients fCBT présentait un PHQ-9 baseline plus bas que les patients iCBT. Pourquoi est-ce problématique et comment les auteurs l'abordent-ils ?

Réponse détaillée

Le PHQ-9 initial est plus bas en fCBT (11, 57) qu'en iCBT (13, 31). Ce simple fait montre que les groupes ne sont pas comparables sur une dimension clinique centrale. Cela peut refléter une **confusion par indication** : peut-être que les patients plus sévères ont été plus volontiers orientés vers iCBT, ou inversement que certaines logiques d'organisation régionales ont orienté différemment les patients.

Ce déséquilibre est problématique parce qu'il peut fausser l'interprétation du changement. Un groupe qui part plus haut a parfois plus de marge de baisse, mais il est aussi exposé à plus de régression vers la moyenne. Les auteurs abordent ce problème par l'ajustement causal et en testant des analyses incluant explicitement le PHQ-9 de base. Le fait que les résultats changent dans ces sensibilités montre justement que cette variable est importante.

La bonne copie doit dire deux choses : **déséquilibre réel, et réponse analytique sérieuse mais non définitive.**

Réponse type courte

Le déséquilibre de PHQ-9 initial suggère une non-comparabilité des groupes et une possible confusion par indication. Les auteurs essaient de l'atténuer par l'ajustement et les sensibilités, mais cela reste un point de fragilité important.

Question 23

Expliquez le concept de biais écologique et pourquoi il est absent de cette étude.

Réponse détaillée

Le biais écologique apparaît quand on attribue à des individus des relations qui n'ont été observées qu'au niveau agrégé, par exemple des régions, des écoles ou des pays. Ici, ce n'est pas le cas : les auteurs travaillent sur des **patients individuellement identifiés**, avec des covariables individuelles et un outcome individuel (PHQ-9).

Il y a certes une composante géographique, car fCBT est surtout issue d'Uusimaa, mais l'analyse ne repose pas sur une comparaison purement agrégée entre régions. Elle reste à l'échelle individuelle. Le problème principal n'est donc pas l'écologie au sens strict, mais la **confusion structurelle liée au territoire et à l'organisation de l'offre.**

La bonne réponse doit donc refuser le mauvais mot tout en disant quel est le vrai problème.

Réponse type courte

Le biais écologique est absent au sens strict, car l'étude travaille sur des données individuelles. En revanche, elle reste exposée à une confusion régionale et organisationnelle liée à la concentration de fCBT en Uusimaa.

Question 24

Les facteurs génétiques pourraient être des confondeurs de l'association traitement-résultat.
Qu'en pensez-vous ?

Réponse détaillée

Oui, c'est plausible, même si ce n'est probablement pas le premier biais auquel on pense. Des vulnérabilités familiales ou génétiques peuvent influencer la sévérité, la chronicité, les comorbidités, voire certaines caractéristiques comportementales qui modifient l'adéquation à un soin numérique ou présentiel. Si ces dimensions ne sont pas observées directement, elles peuvent alimenter une partie de la confusion résiduelle.

Le papier dispose toutefois d'un grand nombre de variables familiales et diagnostiques antérieures, ce qui réduit partiellement cette inquiétude. Mais il ne mesure pas directement le substrat génétique ou des traits psychologiques fins. La bonne conclusion est donc : **biais possible, probablement indirect, pas central mais pas absurde non plus.**

Réponse type courte

Des facteurs familiaux ou génétiques peuvent contribuer indirectement à la confusion s'ils influencent à la fois le profil clinique et l'adéquation au traitement. L'article ajuste beaucoup, mais ne peut pas exclure totalement cette source de confusion résiduelle.

Question 25

Dans la discussion, les auteurs mentionnent que "digital affinity" pourrait être un facteur important. Pourquoi ?

Réponse détaillée

Parce que l'affinité numérique peut agir comme un **confondeur non mesuré très plausible**. Un patient plus à l'aise avec les outils numériques est probablement plus susceptible :

- d'être orienté vers iCBT ;
- d'accepter cette modalité ;
- de compléter les séances et les questionnaires ;
- d'en tirer un meilleur bénéfice pratique.

Autrement dit, cette variable peut influencer à la fois l'exposition et l'issue. Si elle n'est pas observée, le TMLE ne peut pas la corriger. C'est un exemple parfait de limite substantielle qui reste après tous les raffinements statistiques.

La bonne copie doit citer ce point précisément, parce qu'il est concret et directement lié au mécanisme de l'étude.

Réponse type courte

L'affinité numérique peut orienter le choix vers iCBT et améliorer indépendamment l'adhésion et le résultat. Comme elle n'est pas mesurée, elle constitue un confondeur non observé particulièrement crédible.

Thème 6. Mesures et variables

Question 26

Qu'est-ce que le PHQ-9 ? Justifiez son choix comme critère de jugement principal dans cette étude.

Réponse détaillée

Le PHQ-9 est une échelle brève, validée, largement utilisée pour mesurer la sévérité des symptômes dépressifs. Il est particulièrement bien adapté à une étude sur registres car il est :

- simple à administrer ;
- disponible en routine ;
- reproductible ;
- facile à suivre dans le temps.

Dans cet article, ce choix est cohérent avec l'objectif d'analyser des soins réels à grande échelle. Utiliser un critère plus lourd, plus clinique ou plus multidimensionnel aurait probablement réduit la faisabilité du projet. La contrepartie est que le PHQ-9 reste un **self-report symptomatique**. Il ne résume ni la qualité de vie, ni le fonctionnement social, ni la durabilité de l'amélioration.

La bonne réponse doit donc dire : **très bon choix pragmatique, mais mesure incomplète du bénéfice clinique global.**

Réponse type courte

Le PHQ-9 est un critère pragmatique, validé et disponible en routine, donc très adapté à un registre naturaliste. Sa limite est de ne refléter qu'une partie du bénéfice clinique, centrée sur les symptômes auto-rapportés.

Question 27

Le critère de jugement est le changement PHQ-9 (variable continue). Discutez l'intérêt d'utiliser une variable continue plutôt qu'un critère dichotomique.

Réponse détaillée

Une variable continue conserve davantage d'information qu'un seuil binaire du type "répondeur / non répondeur". Ici, cela permet de tirer parti de toute la distribution des changements de PHQ-9 et d'améliorer la précision statistique. Dans une étude déjà confrontée à beaucoup d'hétérogénéité, c'est un choix méthodologiquement pertinent.

Un critère dichotomique aurait été plus intuitif pour certains cliniciens, mais au prix d'une perte d'information et d'une dépendance à un seuil arbitraire. La bonne stratégie est souvent celle adoptée implicitement ici : garder la variable continue pour l'analyse principale, puis discuter secondairement le sens clinique.

En examen, il faut aussi rappeler que le caractère continu du critère rend cohérente l'utilisation d'un ATE en points de PHQ-9, ce qui est plus fin qu'une simple proportion de "rémission".

Réponse type courte

Le choix d'une variable continue évite la perte d'information liée à une dichotomisation et améliore la précision statistique. En contrepartie, l'interprétation clinique est parfois moins immédiate qu'avec un seuil de réponse ou de rémission.

Question 28

Le score OASIS est utilisé comme prédicteur. Qu'apporte-t-il dans l'analyse ?

Réponse détaillée

OASIS mesure la sévérité anxieuse et son retentissement. Dans ce papier, il apporte une information pronostique importante, car l'anxiété comorbide peut modifier à la fois la réponse au traitement et le type de modalité vers laquelle un patient est orienté.

Son inclusion renforce donc l'ajustement. Elle permet aussi de mieux modéliser les différences de profil initial entre patients. C'est particulièrement utile dans un contexte où l'on cherche à reconstituer une comparaison contrefactuelle crédible.

La bonne réponse n'est pas seulement "c'est une covariable de plus", mais "c'est une covariable cliniquement pertinente qui peut affecter simultanément exposition et outcome".

Réponse type courte

OASIS apporte une information utile sur l'anxiété comorbide, qui peut influencer à la fois l'orientation thérapeutique et l'évolution clinique. Son inclusion améliore donc la qualité de l'ajustement causal.

Thème 7. Éthique et réglementaire

Question 29

Les auteurs indiquent que le consentement éclairé n'était pas requis pour cette étude. Qu'en pensez-vous ?

Réponse détaillée

Dans une étude rétrospective sur registres pseudonymisés, sans intervention additionnelle imposée aux patients, l'absence de consentement individuel peut être réglementairement acceptable. Le papier précise d'ailleurs le cadre institutionnel et l'autorisation éthique, ce qui est important.

Mais une bonne copie ne se contente pas de dire "autorisé donc sans problème". Il faut rappeler que cela impose une gouvernance rigoureuse des données, une justification claire de l'usage secondaire, et une vigilance sur la confidentialité. Le fait que quatre personnes aient objecté à l'usage de leurs données rappelle que ce sujet n'est pas purement théorique.

La conclusion attendue est donc nuancée : **acceptable dans ce cadre, mais seulement parce que la gouvernance et la proportionnalité sont explicitement prises au sérieux.**

Réponse type courte

L'absence de consentement individuel peut être justifiée dans une étude rétrospective sur registres pseudonymisés, à condition qu'il existe un encadrement éthique et réglementaire solide. C'est acceptable ici, mais ce n'est jamais un détail anodin.

Question 30

Les auteurs précisent qu'il n'y a pas eu d'implication de personnes ayant une expérience vécue dans la conception de l'étude. Que penser de ce point ?

Réponse détaillée

Ce n'est pas une faute méthodologique majeure au sens strict, mais c'est une limite importante de pertinence clinique et sociale. Des usagers auraient pu aider à choisir des outcomes plus proches des priorités vécues, à mieux interpréter le sens d'une différence faible de PHQ-9, ou à mieux réfléchir au dropout dans les parcours réels.

Dans un domaine comme la santé mentale, cette implication devient de plus en plus attendue. Une bonne copie doit donc montrer qu'elle sait élargir l'évaluation au-delà de la seule technique statistique.

Réponse type courte

L'absence d'implication d'usagers n'invalide pas le papier, mais elle limite sa pertinence centrée patient. Elle aurait pu enrichir le choix des outcomes et l'interprétation clinique des résultats.

Question 31

Le financement de l'étude est public. En quoi est-ce important ?

Réponse détaillée

Un financement public réduit le soupçon de conflit d'intérêts commercial direct. Dans ce type d'étude non médicamenteuse, c'est un point rassurant, car il limite l'idée que la question ou la présentation des résultats soient pilotées par un intérêt industriel.

Cela ne protège évidemment pas contre tous les biais : les biais méthodologiques, institutionnels ou académiques existent toujours. Mais ce détail reste important pour le jugement global de crédibilité et d'indépendance du travail.

Réponse type courte

Le financement public réduit les soupçons de conflit d'intérêts commercial direct et renforce l'indépendance perçue du travail, sans supprimer les autres sources possibles de biais.

Thème 8. Santé publique et clinique

Question 32

Si vous étiez décideur de santé publique en France ou au Liban, comment utiliseriez-vous les résultats de cette étude ?

Réponse détaillée

Je n'utiliserais pas cet article pour déclarer que iCBT doit remplacer partout fCBT. En revanche, je l'utiliserais comme argument fort en faveur d'une **offre de soins graduée**, dans laquelle iCBT constitue une modalité accessible, scalable, potentiellement efficace, et particulièrement utile quand l'offre présente est insuffisante.

L'article suggère qu'en moyenne, iCBT n'est pas moins bonne et pourrait même être légèrement meilleure dans ce système de soins. Cela peut aider à légitimer un investissement dans les plateformes numériques, la formation, et les parcours combinés. Mais la transposition doit rester prudente, car le système finlandais, la culture numérique et l'organisation des soins ne sont pas universels.

La bonne réponse mêle donc **prudence de généralisation** et **utilité pratique pour la décision**.

Réponse type courte

J'utiliserais ce papier pour soutenir le développement d'une offre de soins graduée incluant iCBT, surtout en contexte de pénurie d'accès au présentiel. En revanche, je ne l'utiliserais pas comme justification pour remplacer uniformément fCBT par iCBT.

Question 33

Discutez le modèle de soins "stepped care" mentionné dans la discussion. En quoi l'iCBT peut-il y trouver sa place ?

Réponse détaillée

Le stepped care consiste à commencer par des interventions plus accessibles, moins intensives ou moins coûteuses, puis à intensifier si la réponse est insuffisante. iCBT s'intègre très bien dans cette logique : elle peut constituer une première marche pour des patients compatibles avec ce format, tout en laissant la possibilité d'escalader vers fCBT ou vers des prises en charge plus spécialisées.

Ce papier est utile pour ce raisonnement, car il suggère que iCBT peut produire des résultats au moins comparables dans un système réel, sans nécessiter le même niveau de ressources qu'un suivi présentiel généralisé. La bonne copie doit cependant rappeler que cela suppose une bonne sélection des patients et une vraie possibilité de montée en intensité quand nécessaire.

Réponse type courte

iCBT peut s'intégrer naturellement dans un modèle de stepped care comme modalité initiale ou intermédiaire, à condition de prévoir une escalade vers des soins plus intensifs chez les patients qui ne répondent pas suffisamment.

Thème 9. Critique globale

Question 34

Que pensez-vous de la conclusion de l'article telle que présentée dans l'abstract ?

Réponse détaillée

La conclusion de l'abstract est globalement cohérente avec les résultats : le signal principal est en faveur de iCBT, et plusieurs analyses de sensibilité convergent dans la même direction. Mais la formulation peut faire oublier des limites importantes :

- étude observationnelle ;
- non-pré-enregistrement ;
- déséquilibre très fort entre groupes ;
- données manquantes non triviales ;
- variables potentiellement importantes non mesurées.

L'abstract est donc acceptable, mais il doit être relu avec l'esprit critique adéquat. En examen, il faut montrer qu'on sait lire une conclusion à travers la qualité réelle du design.

Réponse type courte

La conclusion est défendable, mais elle doit être lue avec plus de prudence qu'un résultat d'essai randomisé. Le signal est intéressant, sans pour autant effacer les limites structurelles du design observationnel.

Question 35

Quelle est, selon vous, la principale force méthodologique et la principale limite de l'étude ?

Réponse détaillée

La principale force est la combinaison rare d'un **très grand jeu de données naturalistes** et de méthodes causalistes modernes sérieusement déployées. Peu d'articles réussissent à articuler ainsi richesse des registres, question clinique concrète et appareil statistique aussi sophistiqué.

La principale limite reste l'absence de randomisation et donc la possibilité persistante d'une **confusion non mesurée**. Cette limite n'est pas abstraite : elle se matérialise très probablement autour du choix de la modalité thérapeutique, de l'affinité numérique, des préférences du patient, et de l'organisation régionale des soins.

La meilleure conclusion d'examen tient donc les deux extrêmes ensemble : **papier méthodologiquement fort, mais pas preuve causale définitive**.

Réponse type courte

La force majeure est l'utilisation ambitieuse de données naturalistes riches avec des méthodes causalistes modernes. La limite majeure est la confusion non mesurée qui subsiste inévitablement dans une étude sans randomisation.

4. Réponses prêtes à l'emploi

4.1. Quand il faut discuter la précision plutôt que la puissance

Dans cette étude observationnelle, la bonne discussion porte d'abord sur la précision de l'estimation et sur la plausibilité causale. Si l'on force néanmoins un calcul de puissance a posteriori à partir de l'IC95%, on retrouve seulement un ordre de grandeur, qui ne remplace pas l'analyse de validité interne.

4.2. Quand il faut discuter une grande cohorte sans se faire piéger

La grande taille de l'échantillon améliore la précision statistique, mais elle ne garantit pas la comparabilité des groupes. Ici, la vraie question est moins "y a-t-il assez de sujets?" que "les groupes sont-ils suffisamment comparables pour soutenir une interprétation causale?".

4.3. Quand il faut parler du TMLE sans réciter

Le TMLE est un estimateur causal moderne qui combine modèle de traitement et modèle d'outcome pour cibler l'ATE. Il exploite mieux les données observées qu'une régression simple, mais il ne corrige pas les confondeurs non mesurés.

5. Checklist finale

- Ai-je bien identifié le design comme observationnel naturaliste ?
- Ai-je distingué différence brute et ATE ?
- Ai-je expliqué au moins une hypothèse causale fragile ?
- Ai-je relié ma réponse à un chiffre concret du papier ?
- Ai-je distingué signal statistique et ampleur clinique ?
- Ai-je terminé par une conclusion prudente et utilisable à l'examen ?

6. Conclusion

Le sujet EPI 2026 est un très bon test d'examen parce qu'il oblige à ne pas se laisser impressionner par la sophistication technique seule.

Trois idées doivent rester en tête :

1. **Le papier est méthodologiquement fort, mais il reste observationnel.**
2. **Le signal statistique existe, mais l'effet clinique moyen reste modeste.**
3. **La vraie difficulté est d'articuler précision, causalité et prudence interprétative.**

Si vous gardez cette triple ligne de force en tête, vous pourrez répondre proprement à l'ensemble des questions du sujet.