

Comment interpréter un essai clinique négatif?

Prof. Angèle Gayet-Ageron, *Epidémiologie clinique*

L'oratrice ne déclare pas de conflit d'intérêt avec l'industrie mais précise qu'elle n'est pas neutre car impliquée dans l'étude du jour.

Pour rappel, l'efficacité d'un nouveau traitement est évaluée dans une étude randomisée contrôlée de phase 3 lorsque des signaux positifs sont perçus en phase 2. Il est important de rester neutre entre les traitements testés.

Deux hypothèses sont alors possibles:

- Hypothèse nulle (H_0): efficacité nouveau traitement = efficacité du placebo/référence
- Hypothèse alterne (H_A): efficacité nouveau traitement \neq efficacité placebo/référence

Il faut alors calculer le nombre de sujets nécessaires, pour atteindre un certain nombre de quotas, dont la puissance statistique qui est de 80% par convention, ainsi que la différence minimum à détecter, qui doit être définie par un consensus d'expert pour une étude donnée.

Conduire l'essai: recrutement, collecte de données, puis analyses statistiques..

Le traitement est approuvé lorsque le bénéfice de celui-ci est statistiquement significatif...mais comment interpréter si ce n'est pas significatif? Deux réponses possibles:

- manque de puissance \rightarrow besoin d'un nouvel essai avec plus de patients?
- le traitement n'est pas efficace

Certains ont proposé d'interpréter l'évidence selon la valeur P.... Il est important de rappeler la phrase de Carl Sagan: "l'absence de la preuve n'est pas la preuve de l'absence."

La question de recherche de [l'article du jour](#) est donc:

"Quelle est la force de la preuve en faveur de H_0 plutôt que H_A dans les essais cliniques randomisés et contrôlés avec un résultat primaire négatif? (négatif = non significatif)

P-VALUE	INTERPRETATION
0.001	HIGHLY SIGNIFICANT
0.01	
0.02	
0.03	
0.04	SIGNIFICANT
0.049	
0.050	OH CRAP REDO CALCULATIONS.
0.051	ON THE EDGE OF SIGNIFICANCE
0.06	
0.07	HIGHLY SUGGESTIVE, SIGNIFICANT AT THE P<0.10 LEVEL
0.08	
0.09	
0.099	HEY LOOK AT THIS INTERESTING SUBGROUP ANALYSIS
≥ 0.1	

Le démarche diagnostique est similaire à l'interprétation d'un essai clinique.

Il y a une probabilité de maladie à priori, qui est combinée à un examen et donne une probabilité post-test de maladie.

L'examen apporte un rapport de vraisemblance RV, soit la probabilité du résultat sachant qu'on est malade, sur la probabilité du résultat sachant qu'on est non malade.

Rapport de vraisemblance (RV)

$$\frac{P(T|M +)}{P(T|M -)}$$

Cette analogie peut être appliquée à l'interprétation d'un essai, soit: probabilité du résultat négatif, sous l'hypothèse nulle, par rapport à la probabilité du résultat négatif sous l'hypothèse alterne.

Rapport de vraisemblance (RV)

$$\frac{P(\text{Résultat}|H_0)}{P(\text{Résultat}|H_A)}$$

Si

- = 1 : Les 2 hypothèses se valent
- > 1 : H₀ prédit mieux les résultats que H_A
- < 1 : H_A prédit mieux les résultats que H₀

L'oratrice et ses collègues mettent en place une étude transversale, en prenant toutes les études randomisées, contrôlées et négatives, publiées dans 6 journaux prestigieux en 2021.

Inclusion: RCT de phase 3 et de supériorité, résultat négatif sur l'issue primaire ou co-primaire, intervalle de confiance et valeur P disponibles, tout comme la différence à détecter utilisée pour calculer la taille de l'échantillon nécessaire (H_A).

Ces différents éléments permettent de calculer le rapport de vraisemblance de H₀/H_A. Une application pratique est disponible sur [cette page](#).

Résultats

445 essais sont filtrés, et 130 sont identifiés comme négatifs. 169 résultats sont analysés car certains articles ont des co-issues primaires...

La majorité des essais négatifs proviennent du JAMA, ont deux groupes, sont médicamenteux. Dans > ¼ des cas, la comparaison est un placebo ou un sham placebo.

Ce tableau montre les résultats des calculs de vraisemblance des 169 études analysées.

Pour rappel, un RV supérieur à 10 est une preuve forte en faveur de l'hypothèse nulle par rapport à l'alterne dans cette situation.

Au-dessus de 100, c'est une preuve décisive.

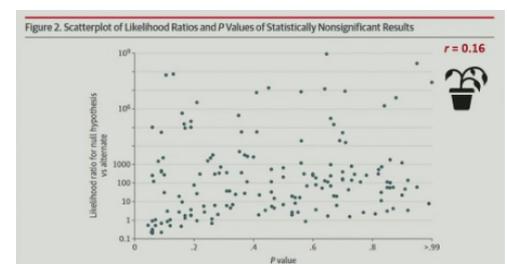
Près de 70% des ces RV (likelihood ratio) calculés sont > 10 (en faveur de H₀), et plus de 50% sont > 100, soit une preuve décisive en faveur de H₀. 30% ont un rapport supérieur à mille.

Table 3. Distribution of Likelihood Ratios for the Null Hypothesis vs the Alternate Hypothesis

Likelihood ratio ^a	No. (%) (n = 169)
>0.1 to 1	15 (8.9)
>1 to 3	18 (10.7)
>3 to 10	19 (11.2)
>10 to 30	13 (7.7)
>30 to 100	16 (9.5)
>100 to 300	20 (11.8)
>300 to 1000	18 (10.7)
≥1000	50 (29.6)

Les auteurs s'intéressent également à la corrélation entre le rapport de vraisemblance et la valeur P, dans l'idée de déterminer si la valeur P peut aider à l'interprétation des études négatives.

Malheureusement, la corrélation est extrêmement faible P. Interpréter la valeur P n'est donc que peu informatif. →



Faiblesses de l'étude

- Pas de données strictes sur la prévalence des essais cliniques négatifs.
- Biais de sélection: seuls 6 journaux sur une seule année.
- La différence à détecter est un choix arbitraire des auteurs de ces études analysées, qui est souvent influencé par la faisabilité de l'étude, car une grande différence à détecter aboutit à moins de sujets dans l'étude.
- Seules deux hypothèses sont utilisées, alors qu'un panel de valeurs possibles

Forces de l'étude

- Analyse au-delà de " le traitement n'a pas amélioré l'issue de façon significative"
- Permet de clarifier les perspectives d'un sujet d'étude: continuer ou s'arrêter.
- Le calcul est aisé une fois les données récoltées, grâce à l'[application](#).
- Cela renforce l'intérêt d'interpréter l'IC 95% plutôt que la valeur P.
- Renforce la lecture critique des articles publiés avec un résultat négatif.
- Réduit l'ambiguïté concernant l'interprétation des IC95% larges

Message clefs

- Un essai négatif ne permet pas de conclure à l'absence d'efficacité du traitement testé
- si l'intervalle de confiance à 95% est utile pour évaluer la plausibilité de l'efficacité et pour suggérer un manque de puissance,
- Le calcul du rapport de vraisemblance (H0/HA) complète l'interprétation du résultat négatif et aide à la décision

Questions

- Est-ce que cela peut être utilisé pour déterminer si une étude doit être interrompue?
- Pas vraiment non, car lors de l'analyse intermédiaire il y a un manque de puissance problématique.
- Dans une étude stratifiée, le rapport de vraisemblance ne va donner qu'une moyenne qui exclut potentiellement un groupe de patients qui pourraient bénéficier du traitement même si le résultat dans la population globale est en faveur de l'hypothèse nulle.
- C'est l'intérêt des analyses en sous-groupe. Ce n'est effectivement pas si simple et il y a un panel de résultats à intégrer: RV, IC, risques d'effets indésirables, analyses de sous-groupes... sont à mettre en balance avant une décision finale.



Compte-rendu de Valentine Borcic
valentine.borcic@gmail.com
Transmis par le laboratoire MGD
colloque@labomgd.ch