

Équivalence et Non-infériorité

Éléments d'Introduction

Lionel RIOU FRANÇA

INSERM U669

Mars 2009

Exemple d'Introduction

- Données tirées de Brinkhaus B et al. Arch Intern Med. 2006;166:450-57.
- Évaluation de l'acupuncture dans le traitement de la lombalgie :
 - Acupuncture,
 - Acupuncture « minimale » (aiguilles en dehors des points d'acupuncture),
 - « Liste d'attente » (pas d'acupuncture avant 8 semaines).
- Critère de jugement : intensité de la douleur (VAS).

Taille d'Échantillon

« The study was powered to detect a group difference of 10 mm in the main outcome measure with 80% power, assuming an SD of 22.5 mm in the primary outcome in both groups and a 2-sided significance level of 5%. »

$$n = \left(Z_{1-\frac{\alpha}{2}} + Z_{1-\beta} \right)^2 \frac{\sigma_0^2 + \sigma_1^2}{(\mu_0 - \mu_1)^2} = (1.96 + 0.84)^2 \frac{2 * 22.5^2}{10^2} \approx 79$$

N.B. Les auteurs optent pour un schéma de randomisation 2:1:1 (!), on a au final 140 acupunctures, 70 minimaux et 74 témoins évalués (resp. 147, 75 et 79 randomisés).

Décision Classique

- 2 hypothèses en compétition :
 - $H_0 : \mu_0 = \mu_1$ (pas de différence entre les groupes),
 - $H_1 : \mu_0 \neq \mu_1$ (les deux groupes diffèrent).
- 2 risques d'erreur :
 - Accepter H_1 alors que H_0 est vraie (erreur α de type I),
 - Accepter H_0 alors que H_1 est vraie (erreur β de type II – $1 - \beta =$ puissance du test),
- Décision :
 - Calcul de p , la probabilité d'observer une différence au moins aussi importante que celle de l'essai, si H_0 est vraie,
 - $p < \alpha \Rightarrow$ rejet de H_0 .

Résultats – 8^e semaine

Critère de jugement : évolution de la douleur depuis la randomisation ($D_{t_0} - D_{t_8}$).

- Témoins : $\hat{\mu}_0 = 6.9$, $\hat{\sigma}_0 = s_0 = 22.0$, $n_0 = 74$,
- Minimalistes : $\hat{\mu}_1 = 23.6$, $\hat{\sigma}_1 = s_1 = 31.0$, $n_1 = 70$,
- Acupuncture : $\hat{\mu}_2 = 28.7$, $\hat{\sigma}_2 = s_2 = 30.3$, $n_2 = 140$.

Test t de Student :

$$t = \frac{\mu_0 - \mu_1}{\sqrt{\frac{(n_0-1)s_0^2 + (n_1-1)s_1^2}{n_0+n_1-2}} \sqrt{\frac{1}{n_0} + \frac{1}{n_1}}}$$

- Acupuncture vs. Témoins : $t = -5.47$, $p < 0.01$,
- Acupuncture vs. Minimalistes : $t = -1.14$, $p = 0.26$.

N.B. Tests multiples !

Les Essais de Supériorité

- Gold Standard : Essai contre Placebo. Contrôle :
 - l'évolution spontanée de la maladie,
 - la régression vers la moyenne,
 - l'effet placebo.
- Permet, avec une taille d'échantillon minimale (différentiel d'efficacité maximisé), de démontrer la supériorité de l'intervention testée.
- Mais problèmes éthiques et pratiques :
 - Pertes de chance si traitement efficace existe,
 - Recrutement difficile si le patient sait qu'il risque d'être dans un bras placebo,
 - Arrêts de traitement plus fréquents si inefficacité perçue.

Alternatives et Nomenclature

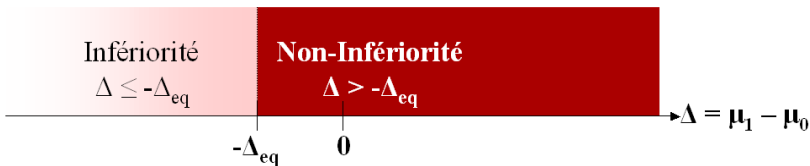
- Essai de *supériorité* : B est plus efficace que A.
- Essai d'*équivalence* : B et A sont thérapeutiquement équivalents.
- Essai de *non-infériorité* : B n'est pas plus mauvais que A.
- L'essai de supériorité demande de définir une différence minimale d'efficacité Δ_{sup} que l'on veut mettre en évidence avec une puissance suffisante (calcul de taille d'échantillon).
- L'essai d'équivalence demande de définir une *marge d'équivalence* à partir d'une grandeur Δ_{eq} qui correspond à la plus grande perte d'efficacité que l'on peut tolérer pour conclure que l'un des traitements n'est pas inférieur à l'autre. $\Delta_{eq} < \Delta_{sup}$.

Justification d'un Essai de Non-Infériorité/Équivalence

- Il existe déjà un traitement comparateur ayant fait la preuve de son efficacité contre placebo,
- L'innovation n'est pas supérieure en efficacité, mais apporte d'autres avantages (tolérance, coût, facilité d'utilisation. . .).

Essai de Non-Infériorité

- On veut montrer que les deux traitements sont comparables, mais on accepte que le traitement fasse un peu moins bien que la référence : $\Delta = \mu_1 - \mu_0 > -\Delta_{eq}$.
- $H_0 : \Delta \leq -\Delta_{eq}$ (traitement inférieur à référence).
- $H_1 : \Delta > -\Delta_{eq}$ (non-infériorité).

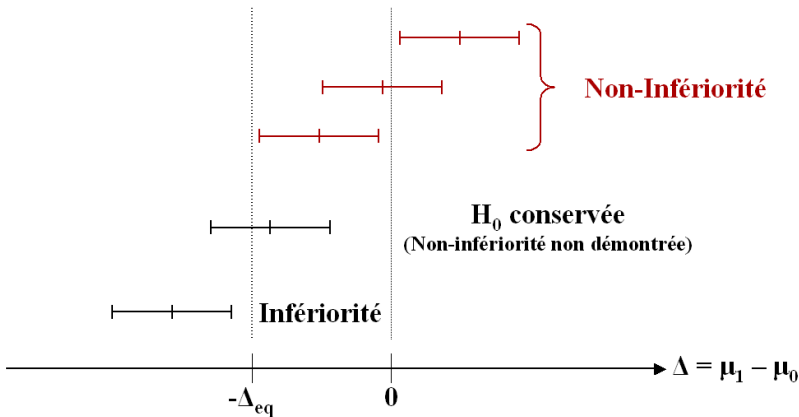


Démonstration de la Non-Infériorité

- Procédure de test (unilatéral) :
 - $\Delta = \mu_1 - \mu_0$.
 - $t = \frac{\hat{\Delta}}{\sqrt{s^2\left(\frac{1}{n_0} + \frac{1}{n_1}\right)}}$ suit une loi de Student décentrée à paramètre de décentrage $h = \frac{-\Delta_{eq}}{\sqrt{s^2\left(\frac{1}{n_0} + \frac{1}{n_1}\right)}}$.
 - $t' = t - h$ suit une loi de Student centrée à $(n_1 + n_0 - 2)$ ddl.
- Approche par intervalle de confiance à $1 - \alpha^1$ de Δ :

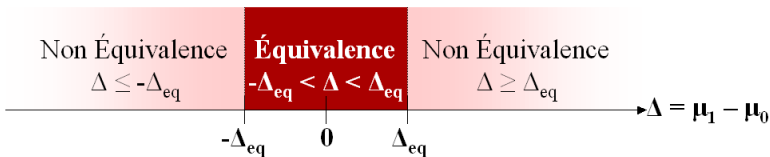
¹Controversé. L'EMA recommande un IC unilatéral à $1 - \frac{\alpha}{2}$ ou un IC bilatéral à $1 - \alpha$, la coutume en bioéquivalence est d'utiliser l'IC bilatéral à $1 - 2\alpha$.

Démonstration de la Non-Infériorité



Essai d'Équivalence

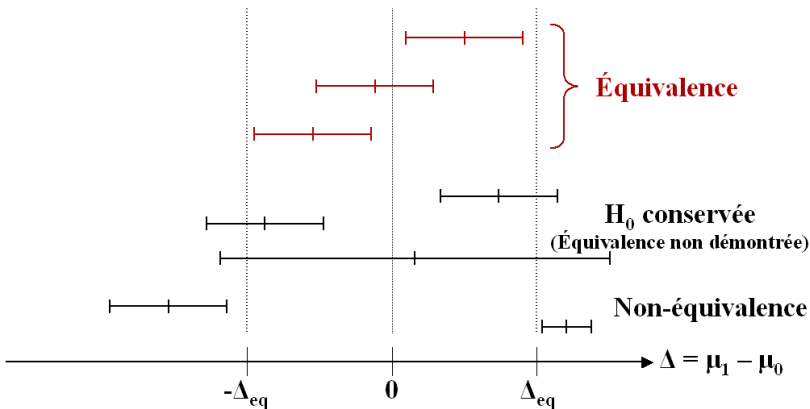
- On veut montrer que $|\Delta| = |\mu_1 - \mu_0| < \Delta_{eq}$.
- $H_0 : \Delta \leq -\Delta_{eq}$ ou $\Delta \geq \Delta_{eq}$ (non équivalence).
- $H_1 : -\Delta_{eq} < \Delta < \Delta_{eq}$ (équivalence).



Démonstration de l'Équivalence

- Procédure de test (bilatéral) :
 - $\Delta = \mu_1 - \mu_0$.
 - 2 hypothèses nulles :
 - $H'_0 : \Delta = -\Delta_{eq}$,
 - $H''_0 : \Delta = \Delta_{eq}$.
 - \Rightarrow 2 paramètres de décentrage h et $-h$ et 2 statistiques de test $t' = t - h$ et $t'' = t + h$.
 - $p = \max(p', p'')$.
- Approche par intervalle de confiance à $1 - 2\alpha$ de Δ :

Démonstration de l'Équivalence



Nombre de Sujets Nécessaire

Le calcul dépend :

- Du risque α de rejeter H_0 à tort,
- Du risque β qui détermine la puissance $1 - \beta$ de montrer l'équivalence/non-infériorité quand elle existe,
- De la différence δ d'efficacité entre les deux traitements (en général, on postule $\delta = 0$),
- De σ_δ ,
- De la marge d'équivalence et donc de Δ_{eq} .

Nombre de Sujets Nécessaire

Dans le cas d'une différence de moyennes,

- Supériorité : $n = \frac{2\sigma^2}{(\Delta_{sup})^2} (z_{1-\frac{\alpha}{2}} + z_{1-\beta})^2$,
- Équivalence : $n = \frac{2\sigma^2}{(\Delta_{eq}-\delta)^2} (z_{1-\alpha} + z_{1-\frac{\beta}{2}})^2$,
- Non-infériorité : $n = \frac{2\sigma^2}{(\Delta_{eq}-\delta)^2} (z_{1-\alpha} + z_{1-\beta})^2$.

Comme $\Delta_{sup} \gg \Delta_{eq}$, les tailles d'échantillon sont plus importantes en non-infériorité qu'en supériorité.

Aspects du Protocole

- Essais très sensibles à la qualité du protocole.
- Calquer au maximum l'essai qui a établi la supériorité du traitement de référence (→ démontrer l'équivalence du nouveau traitement dans le contexte où on sait le traitement de référence efficace).
- Mêmes critères d'inclusion, mêmes patients que dans l'essai de supériorité,
- Même critère de jugement,
- Bon choix du traitement de référence,
- Bon choix de la marge d'équivalence.

Choix du Traitement de Référence

- Efficacité établie, contre placebo (non-inférieur à un traitement évalué contre placebo ne suffit pas),
- Mêmes conditions d'usage (dose, administration...) que dans l'essai de supériorité,
- « Assay sensitivity » : capacité de l'étude à distinguer un traitement efficace d'un traitement inefficace.
 - Validation interne (mieux) : bras placebo dans l'essai d'équivalence.
 - Validation externe (moins bon) : montrer que l'efficacité du traitement de référence dans l'essai de non-infériorité est la même que dans l'essai de supériorité.

Population d'Analyse : Recommandations

- Analyse en Intention de Traiter (ITT) : tous les patients randomisés, dans leur groupe de randomisation. Recommandé pour l'essai de supériorité, déconseillé en équivalence.
- Analyse Per Protocole (PP) : patients n'ayant pas eu de changement/arrêt de traitement, ni violation de protocole. Recommandée en équivalence, mais à compléter par une analyse ITT.

Population d'Analyse : Justification

- ITT : déviations au protocole semblables (en principe) dans les 2 groupes \Rightarrow homogénéisation des réponses \Rightarrow diminution de l'efficacité estimée. Souhaitable (conservateur) en supériorité, mauvais en équivalence.
- PP \Rightarrow augmentation du contraste entre les groupes attendue. Souhaitable (conservateur) en équivalence. Mais risque de surestimation de l'efficacité d'un traitement où les non-répondeurs arrêteraient systématiquement (i.e. conclure à la non-infériorité alors qu'il y avait infériorité).

\Rightarrow Les essais d'équivalence/non-infériorité doivent se faire en PP, mais très important de décrire la population PP !

Qui Détermine la Marge d'Équivalence ?

- Bouvenot & Villani : « La méthode consiste à faire définir *par les cliniciens* (et non par les statisticiens) (...), une zone dite d'équivalence, c'est-à-dire un intervalle dans lequel il est raisonnable que se situe la différence d'effet observée entre les deux traitements, si l'on admet qu'ils sont également efficaces. »
- D'Agostino : « To the dismay of some, the statisticians seem to have taken control of this issue. »
- EMEA : « The selection of the non-inferiority margin is based upon a combination of statistical reasoning and clinical judgement. »

Comment Déterminer la Marge d'Équivalence ?

Deux contraintes :

- Montrer que le traitement évalué à une efficacité non nulle (i.e. supériorité par rapport à placebo) → étape statistique – inutile si validation interne (bras placebo),
- Montrer qu'il n'est pas inférieur au traitement de référence → étape clinique : Δ_{eq} doit être « cliniquement négligeable ».

En pratique, énorme variabilité dans les méthodes de détermination de Δ_{eq} et dans les recommandations officielles sur ces méthodes.

Recommandations EMEA (Étape Statistique)

- Méta-analyse → estimation de (R-P), l'effet du traitement de référence (R) vs. Placebo (P).
- L'essai d'équivalence donnera directement (T-R), l'effet du traitement évalué (T) vs. R et, par comparaison indirecte, (T-P).
- $-\Delta_{eq}$: plus petite valeur acceptable pour l'IC de (T-R) qui assure que l'IC de (T-P) ne contienne pas 0.

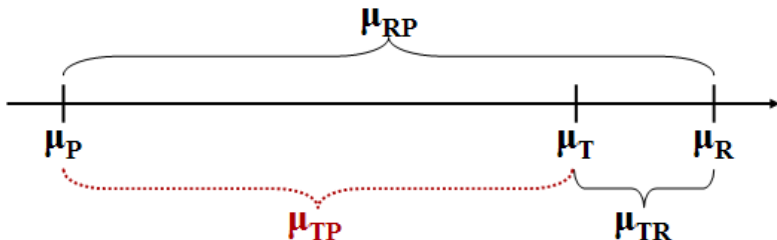
N.B. Hypothèse de « constancy » : l'efficacité du comparateur contre placebo (essais de supériorité) se maintiendra dans l'étude d'équivalence.

Comparaisons Indirectes : Principe

- Soit $\mu_{RP} = \mu_R - \mu_P$ l'effet du traitement R par rapport au traitement P, et $V[\mu_{RP}]$ sa variance,
- Soit $\mu_{TR} = \mu_T - \mu_R$ l'effet du traitement T par rapport au traitement R, et $V[\mu_{TR}]$ sa variance.

Alors :

- $\mu_{TP} = \mu_{TR} + \mu_{RP}$,
- $V[\mu_{TP}] = V[\mu_{TR}] + V[\mu_{RP}]$.



Autres Méthodes pour Déterminer Δ_{eq}

- Prendre la plus petite de toutes les bornes des IC de (R-P),
- T doit conserver $\lambda\%$ de l'efficacité relative de R.
Coutume : $\lambda = 50\%$.
- « This non-inferiority margin cannot be greater than the smallest effect size that the active drug would be reliably expected to have compared with placebo in the setting of a placebo-controlled trial. » (guidelines ICH²).

²International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use

Ah Ben J'm'a Trompé !

- Passage non-infériorité → supériorité :
 - Non-infériorité démontrée,
 - $\delta > 0$ (par intervalle de confiance, ou par test statistique de supériorité).
- Passage supériorité → non-infériorité : jamais³.

³Certains l'admettent si Δ_{eq} est déterminé a priori et si on argumente que H_0 du test supériorité n'est pas rejeté car manque de puissance et pas de différence entre les traitements.

Exemple d'Introduction

- Données tirées de Leblan D et al. Rev Rhum. 2000;67:634-40.
- Évaluation de l'harpagophyton dans le traitement de la gon/coxarthrose :
 - Diacerhéine 100 mg/j,
 - Harpagophyton 2 610 mg/j.
- Critère de jugement : intensité de la douleur (VAS).

Motivation de la non-infériorité

« Le but de cet essai était de démontrer l'efficacité de l'harpagophyton dans l'arthrose pendant une durée de traitement suffisamment longue, égale à quatre mois. Il n'a pas été inclus de groupe placebo dans cette étude car il est souvent difficile de recruter des patients lorsqu'ils sont informés du risque (égal à 50 %) d'être dans un groupe placebo, et ce d'autant plus que la durée de traitement est particulièrement longue (quatre mois). Il a donc été choisi comme traitement comparateur un anti-arthrosique d'effet différé et prolongé de référence, la diacerhéine, qui a fait l'objet de nombreux essais cliniques, démontrant son efficacité à la fois dans des études contre placebo ou contre anti-inflammatoires non stéroïdiens ».

Taille d'Échantillon

« En posant l'hypothèse que la différence vraie entre les traitements était nulle avec un Δ de 10 mm, un test d'équivalence unilatéral avec un risque α à 0,05, un risque β à 0,10 et un écart type de 18 mm, nécessitait l'inclusion de 56 patients par groupe. »

$$n = \frac{2\sigma^2}{(\Delta_{eq} - \delta)^2} (z_{1-\alpha} + z_{1-\beta})^2 = \frac{2 \cdot 18^2}{10^2} (1,64 + 1,28)^2 = 55,5.$$

« Cent vingt-deux patients ont été randomisés dont 92 ont terminé l'essai conformément au protocole. »

Protocole d'Analyse

« L'hypothèse principale à tester était la suivante : essai d'équivalence unilatérale ou de non infériorité de l'harpagophyton par rapport à la diacérhéine sur la douleur spontanée mesurée à l'aide d'une EVA (100 mm). »

« L'analyse statistique a été réalisée perprotocole et en intention de traiter. Le test d'équivalence a été répété pour ces deux populations. »

Non-infériorité

On mesure l'évolution de la douleur entre l'inclusion et J120. Comme une différence négative indique une *diminution* de la douleur, un δ négatif indique qu'Harpagophyton est plus efficace que la Diacerhéine pour diminuer la douleur. La non-infériorité est rejetée si l'intervalle de confiance coupe la valeur $\Delta_{eq} = +10$.

Analyse	Harpagophyton	Diacerhéine	δ	IC _{90%}
PP	-30,6	-25,5	-5,1	(-13,1 ; 3,0)
ITT	-28,8	-22,1	-6,7	(-14,2 ; 0,9)

N.B. IC à $1 - 2\alpha = 90\%$ employé.

Avantages Comparés

« La différence de la consommation d'AINS durant les quatre mois de traitement était significative entre les deux groupes, au profit du groupe harpagophyton ».

« Le nombre de patients ayant présenté au moins un événement indésirable a été significativement plus faible dans le groupe harpagophyton ».

« Putative Placebo Analysis »

« Par ailleurs, l'appréciation de l'intensité de l'effet placebo dans l'arthrose a récemment été étudiée. Une revue portant sur 457 patients traités par placebo dans le cadre de six essais contrôlés d'anti-arthrosiques symptomatiques d'action lente a montré que sous placebo, la baisse moyenne de la douleur à l'EVA était de l'ordre de 10 à 16 mm [12]. Par conséquent, la baisse observée dans le groupe harpagophyton (-30,6 mm) est nettement supérieure à celle habituellement observée sous placebo. »

Bibliographie

- Bouvenot G, Villani P. Les essais cliniques d'équivalence en rhumatologie. Rev Rhum. 2000;67:569-72.
- Brinkhaus B, Witt CM, Jena S, et al. Acupuncture in patients with chronic low back pain. A randomized controlled trial. Arch Intern Med. 2006;166:450-57.
- D'Agostino RB, Massaro JM, Sullivan LM. Non-inferiority trials: design concepts and issues – the encounters of academic consultants in statistics. Statist. Med. 2003;22:169-86.
- Elie C, De Rycke Y, Jais JP, et al. Aspects méthodologiques et statistiques des essais d'équivalence et de non-infériorité. RESP. 2008;56:267-77.
- EMEA/CHMP. Points to consider on switching between superiority and non-inferiority. July 2000.
- EMEA/CHMP. Guideline on the choice of the non-inferiority margin. July 2005.
- Leblan D, Chantre P, Fournié B. L'harpagophyton dans le traitement de la gonarthrose et de la coxarthrose. Résultats à quatre mois d'une étude prospective multicentrique, contrôlée en double aveugle, versus diacérhéine. Rev Rhum. 2000;67:634-40.