

ESSAIS DE NON INFÉRIORITÉ

Vincent Lévy

Département de recherche clinique

GHUPSSD, Hôpital Avicenne, Université Sorbonne Paris Nord, APHP, Bobigny

UMR-1153 INSERM, Equipe ECSTRRA, Hôpital Saint Louis, Paris



A statistical joke called "Non-Inferiority" drug trial

Exemple 1

Journal of Clinical Oncology®

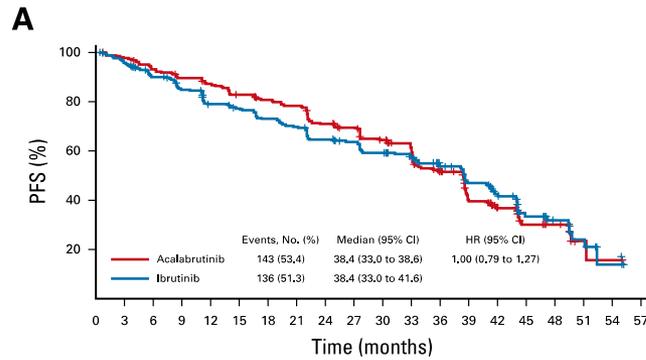
Volume 39, Issue 31 3441

Acalabrutinib Versus Ibrutinib in Previously Treated Chronic Lymphocytic Leukemia: Results of the First Randomized Phase III Trial

John C. Byrd, MD¹; Peter Hillmen, MD, MBChB, PhD²; Paolo Ghia, MD, PhD^{3,4}; Arnon P. Kater, MD, PhD⁵; Asher Chanan-Khan, MD⁶; Richard R. Furman, MD⁷; Susan O'Brien, MD⁸; Mustafa Nuri Yenerel, MD⁹; Arpad Illés, MD¹⁰; Neil Kay, MD¹¹; Jose A. Garcia-Marco, MD, PhD¹²; Anthony Mato, MD¹³; Javier Pinilla-Ibarz, MD, PhD¹⁴; John F. Seymour, PhD¹⁵; Stephane Lepretre, MD^{16,17}; Stephan Stilgenbauer, MD¹⁸; Tadeusz Robak, PhD¹⁹; Wayne Rothbaum, MS²⁰; Raquel Izumi, PhD²⁰; Ahmed Hamdy, MD²⁰; Priti Patel, MD²¹; Kara Higgins, MS²¹; Sophia Sohoni, MD²¹; and Wojciech Jurczak, MD, PhD²²

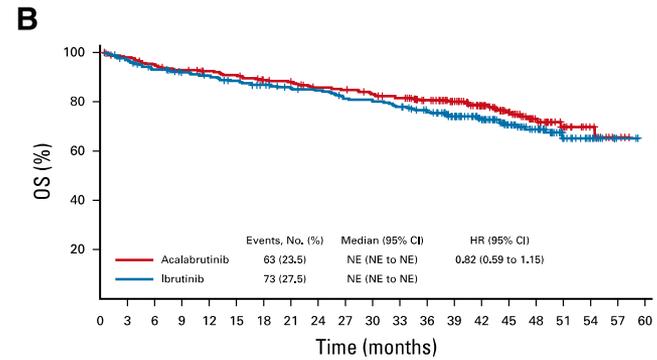
Sample size was calculated assuming a hazard ratio (HR) scale margin of 1.429 (noninferiority margin of 30%) between the acalabrutinib and ibrutinib groups for IRC-assessed median PFS, using a fixed margin method.²⁶

Exemple 1 (suite)



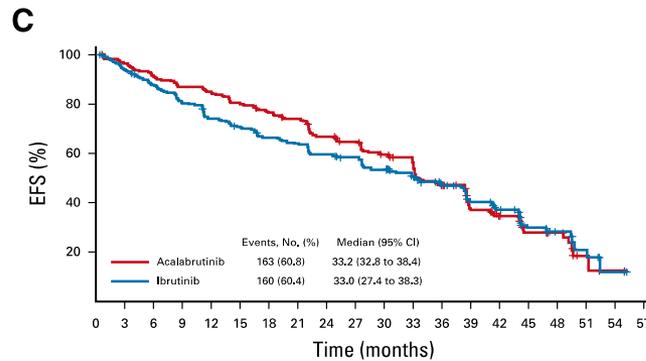
No. at risk:

Time (months)	0	3	6	9	12	15	18	21	24	27	30	33	36	39	42	45	48	51	54	57
Acalabrutinib	268	250	235	227	219	207	200	193	173	163	148	110	84	59	31	21	13	3	1	0
Ibrutinib	265	240	221	205	186	178	168	160	148	142	130	108	81	66	41	26	15	8	2	0



No. at risk:

Time (months)	0	3	6	9	12	15	18	21	24	27	30	33	36	39	42	45	48	51	54	57	60
Acalabrutinib	268	259	247	242	236	231	223	218	210	207	201	196	183	155	127	95	59	32	18	4	0
Ibrutinib	265	252	241	233	227	220	212	205	203	194	191	186	173	143	121	88	60	28	15	2	0



No. at risk:

Time (months)	0	3	6	9	12	15	18	21	24	27	30	33	36	39	42	45	48	51	54	57
Acalabrutinib	268	251	235	228	220	209	199	191	172	162	147	110	85	61	34	23	15	3	1	0
Ibrutinib	265	241	222	205	188	177	167	160	150	142	129	105	79	63	40	25	14	7	2	0

Exemple 2

The NEW ENGLAND JOURNAL of MEDICINE

ORIGINAL ARTICLE

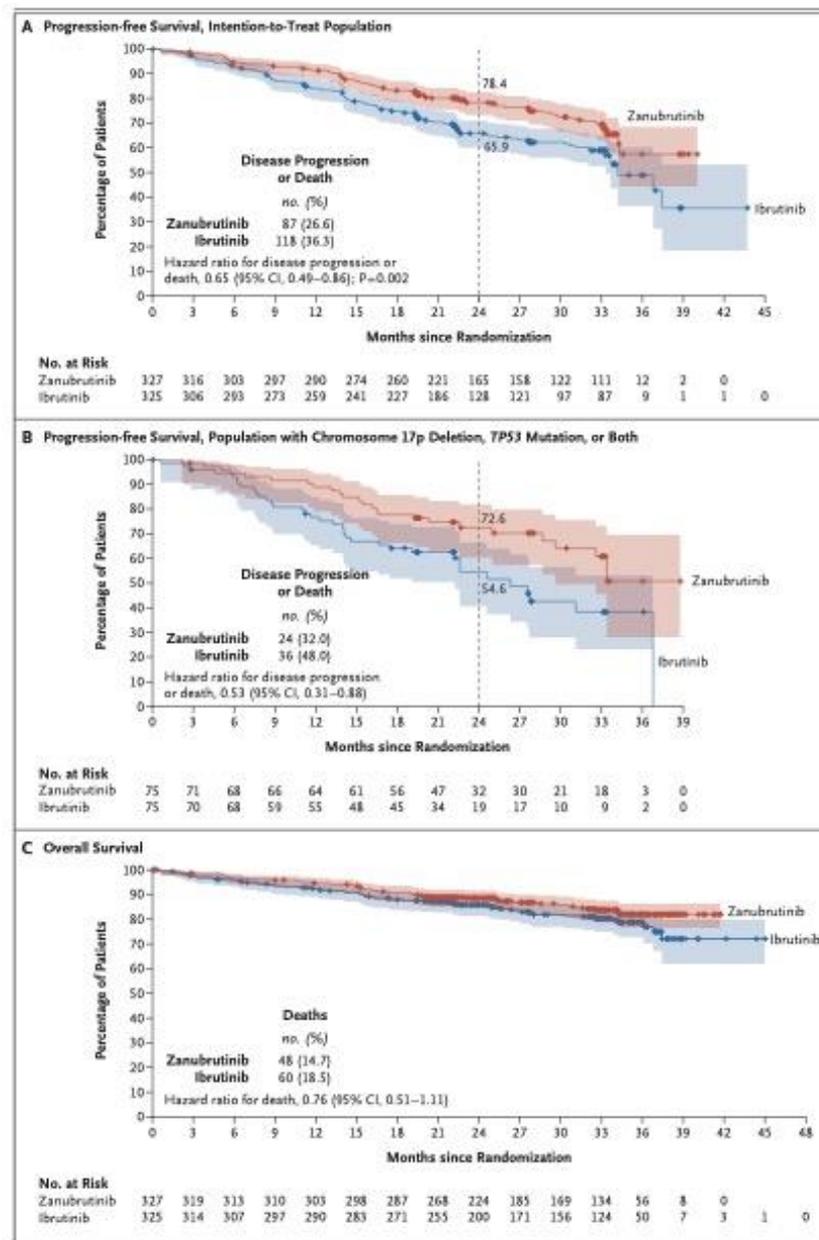
Zanubrutinib or Ibrutinib in Relapsed or Refractory Chronic Lymphocytic Leukemia

J.R. Brown, B. Eichhorst, P. Hillmen, W. Jurczak, M. Kaźmierczak, N. Lamanna, S.M. O'Brien, C.S. Tam, L. Qiu, K. Zhou, M. Simkovic, J. Mayer, A. Gillespie-Twardy, A. Ferrajoli, P.S. Ganly, R. Weinkove, S. Grosicki, A. Mital, T. Robak, A. Osterborg, H.A. Yimer, T. Salmi, M.-D.-Y. Wang, L. Fu, J. Li, K. Wu, A. Cohen, and M. Shadman

We estimated that a sample of 600 patients would provide the trial with greater than 90% power to detect the noninferiority of zanubrutinib to ibrutinib with regard to overall response.

disease progression or death. The noninferiority was tested with a noninferiority margin (hazard ratio) of 1.33 with the use of a stratified Wald

Exemple 2 (suite)



Bibliographie

PUBMED

- RCT : 592 290 essais
- Non inferiority : 2 834 essais
- Non inferiority cancer : 656 essais

Introduction

- Essais d'équivalence : développement dans le cadre des études pharmacologiques de bio-équivalence
 - Hypothèse fondamentale :
mêmes effets biologiques correspondent aux mêmes effets thérapeutiques, en terme d'efficacité et de tolérance
 - Pour formes galénique d'une même molécule
 - Pour molécules de la même classe thérapeutique

Introduction (suite)

- En hématologie (maligne) : « essai de supériorité » ou « essai d'efficacité », permet la comparaison d'un nouveau traitement (**N**) à un placebo (**P**) ou à un traitement de référence (**R**).
- L'essai clinique contrôlé randomisé, en établissant statistiquement la supériorité de l'effet d'un nouveau traitement au bras contrôle, est l'approche expérimentale de référence

Problématique en cas de traitement de référence « efficace »

→ Bénéfices escomptés en terme d'efficacité avec de nouveaux traitements réduits

- Difficulté de la démonstration statistique de leur supériorité

→ Si traitements très efficaces, recherche de nouveaux traitements supérieur

- Soit vain
- Soit nécessiterait d'inclure un très grand nombre de sujets pour pouvoir mettre en évidence une petite différence

Pose la question de la **pertinence clinique** de la différence

- En « dehors » de l'efficacité d'autres critères tels **le coût, la facilité d'utilisation, la voie d'administration, les effets indésirables immédiats ou à long terme** sont à prendre en compte

Qu'est-ce que la non infériorité ?

- Les effets du traitement expérimental / référence **ne portent pas sur l'efficacité** mais peuvent porter sur
 - Diminution de la fréquence/gravité des effets secondaires
 - Facilité d'utilisation, allègement de traitement, prises...
 - Diminution des coûts, intérêt marketing
- Donc : améliorer le **confort, l'observance, l'acceptabilité, la qualité de vie** sans perdre en efficacité

Rappel essai de supériorité

- Très grande majorité des essais dans les maladies malignes
 - Soit ΔS représentant une différence d'effet entre les 2 traitements que l'on veut mettre en évidence
 - ΔS doit être **assez grand** pour correspondre à une différence **cliniquement intéressante** entre les deux traitements et donc pour que l'étude soit pertinente
 - A contrario si ΔS est « trop » grand → étude **irréaliste**
- A l'aide de ΔS on effectue un calcul du nombre de sujets nécessaire pour avoir une puissance suffisante
 - L'analyse consistera essentiellement à **effectuer un test de comparaison sous l'hypothèse nulle d'égalité et à considérer la valeur « p » pour conclure.**

Essais de supériorité

Formulation bilatérale

- $H_0 \quad |\Delta| = |\mu_N - \mu_R| = 0$ N et R ne sont pas \neq
- $H_1 \quad |\Delta| \neq 0$ N et R sont \neq

Formulation unilatérale

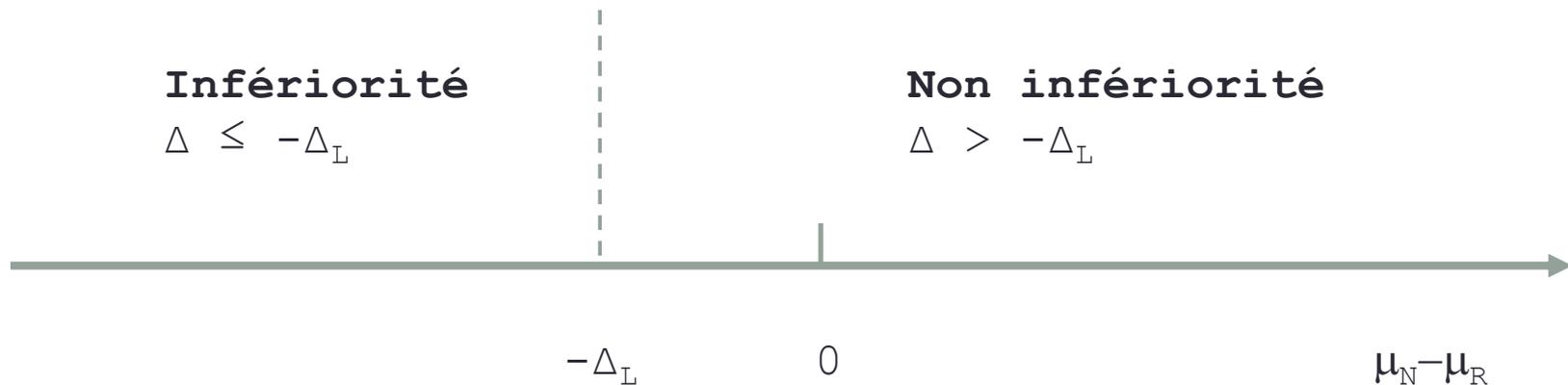
- $H_0 \quad \Delta \leq 0$ N et R ne sont pas \neq
- $H_1 \quad \Delta > 0$ N est supérieur à R

Essai de non infériorité

- On définit Δ_L qui correspond à **la plus grande perte d'efficacité que l'on peut tolérer pour conclure que le nouveau traitement n'est pas inférieur au traitement de référence.**
 - Δ_L doit être suffisamment faible pour ne pas avoir de signification clinique.
 - Donc $\Delta_L < \Delta_S$
- Δ_L est utilisé pour le calcul du nombre de sujets nécessaire pour avoir une puissance suffisante, mais celui-ci interviendra également à l'étape de l'analyse statistique des résultats

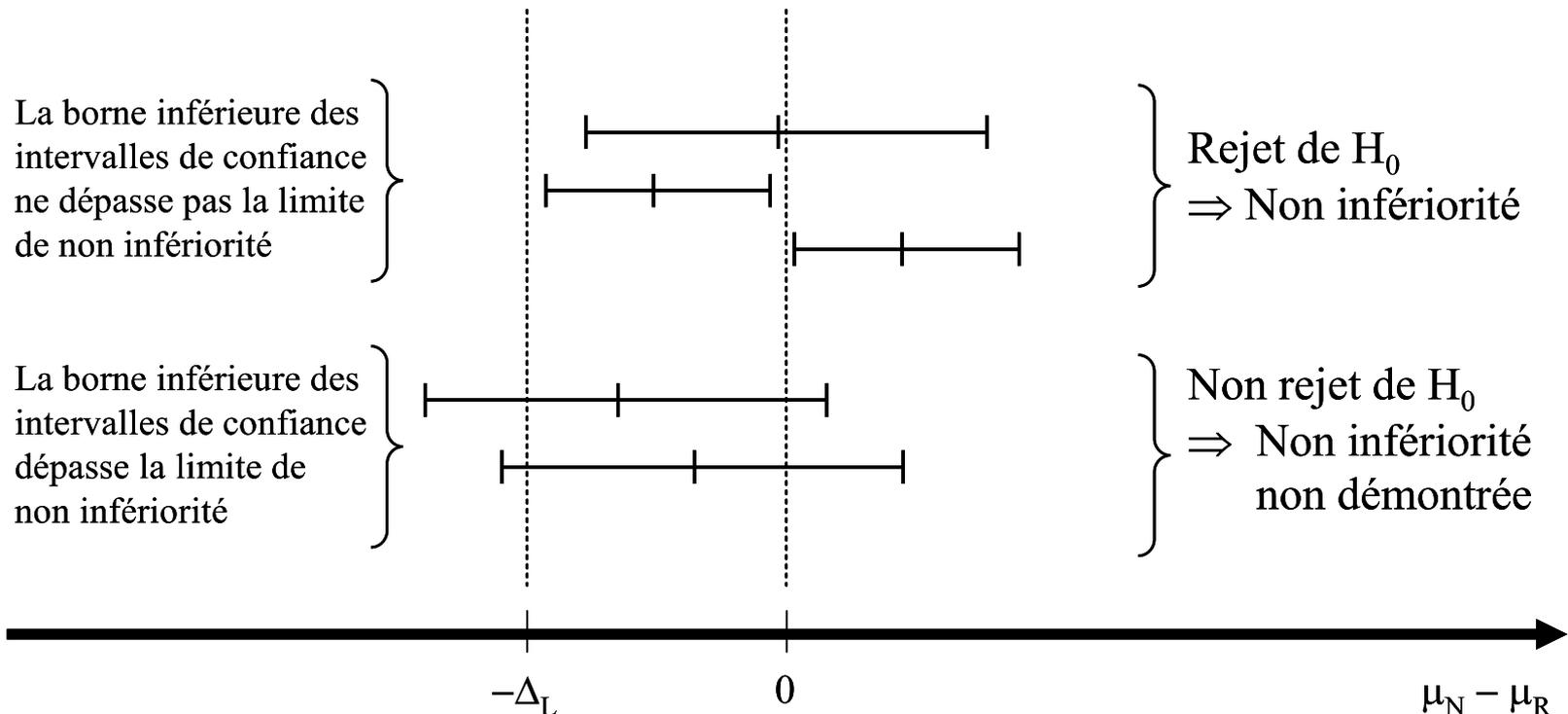
Essai de non infériorité

- On souhaiterait que N soit égal à R ($\mu_N - \mu_R$)
- Mais condition trop forte
- On accepte que N fasse un peu moins bien ($\mu_N < \mu_R$)
- Si N fait mieux que R cela ne nous gêne pas
- Un peu $\Rightarrow -\Delta_L$
- On veut que $\Delta > -\Delta_L$



Essai de non infériorité 3 situations

- Critère **quantitatif** : différence de moyennes
- Critère **binaire** : différence de proportion
- Critère **censuré** : comparaison de courbes de survie



Le protocole

- Objectifs
- Traitement de référence
- Critères d'inclusion
- Critères de jugement
- Borne d'équivalence clinique
- Nombre de sujets nécessaire
- Validation interne et externe
- Population d'analyse

Principale difficulté

- Importance de la méthodologie car les défauts méthodologique tendent le plus souvent à produire des résultats en faveur d'une conclusion d'équivalence
- Donc : méthodologie **RIGOUREUSE**

Objectif

- L'objectif d'équivalence doit être prévue initialement dans le protocole
- Passage supériorité => équivalence/non infériorité possible sous certaines conditions

Traitement de référence (comparateur)

- Efficacité établie
- Administré dans les mêmes conditions que dans les essais ayant démontré son efficacité
- Attention aux traitements concomitants efficaces (*en hématologie supportive care p.e.*)

Critères d'inclusion

- Sujets identiques à ceux sur lesquels le traitement de référence a été démontré efficace

Critère de jugement

- Celui utilisé pour l'évaluation de l'efficacité du traitement de référence
- Mesure correcte de l'efficacité
 - Valide et reconnu dans la pathologie
 - Sensible
 - Sensibilité d'analyse
 - Validation interne ou externe

Validation

- Attention : **conclure à la non infériorité (ou l'équivalent) ne signifie pas que les 2 traitements sont efficaces**
- Validation *interne*
 - Intérêt du bras placebo (rare en hématologie maligne)
- Validation *externe*
 - Comparaison de l'efficacité du traitement de référence avec celle observée dans les études antérieures

Borne d'équivalence (ou de non infériorité)

- Pas de règle stricte
- Doit être « **petite** » (!)
- En général on choisit une fraction de la borne de supériorité
- Importance de la discussion **entre statisticiens et cliniciens+++**
 - Notion de différence cliniquement négligeable
 - Importance d'un consensus d'experts

« The selection of the non-inferiority margin is based upon a combination of statistical reasoning and clinical judgement »

EMA

Calcul du nombre de sujets nécessaires

Les paramètres (1)

- Les paramètres intervenants sont **les mêmes** que pour un essai de supériorité
- Le risque α (probabilité de **rejeter H0 à tort**) :
 - Probabilité de conclure que les deux traitements sont équivalents ou non inférieurs alors qu'en réalité la différence est plus grande que ΔL . On prend en général $\alpha=5\%$ comme en essai de supériorité.
- Le risque β (probabilité de **ne pas rejeter H0 à tort**)
 - Probabilité de ne pas conclure que les deux traitements sont équivalents alors qu'en réalité la différence est plus petite que ΔL . La puissance $1-\beta$ est ici la probabilité de trouver une équivalence ou une non infériorité qui existe effectivement.
 - On choisit le plus souvent une puissance de 80% ou de 90% comme en essai de supériorité.

Les paramètres (suite)

	Essai de supériorité	Essai de non-infériorité
Hypothèse nulle (H_0)	Trait ^{mt} = contrôle	Trait ^{mt} ≤ marge inf
Hypothèse alternative (H_1)	Trait ^{mt} ≠ contrôle	Trait ^{mt} > marge inf
Erreur de type 1	<i>Conclure H_1 alors que H_0 est vrai</i>	
	Conclure Trait ^{mt} ≠ contrôle	Conclure non infériorité alors que infériorité est vrai
Erreur de type 2	<i>Conclure H_0 alors que H_1 est vrai</i>	
	Conclure Trait ^{mt} = contrôle	Conclure infériorité alors que non inférieure est vrai

Calcul du nombre de sujets nécessaires (3)

- Une estimation de la **variabilité du critère de jugement principal**, comme dans un essai de supériorité
- Un terme ΔL
- Une **hypothèse alternative particulière**, définie ici comme l'absence de différence / la non infériorité entre les deux
- La formulation **unilatérale** (non infériorité).

Population d'analyse ITT (1)

- Dans l'analyse en intention de traiter (ITT), le groupe auquel appartient un patient donné au moment de l'analyse est **strictement défini comme étant le groupe dans lequel ce patient a été randomisé**, quels qu'aient été les aléas auxquels le patient a pu être soumis après la randomisation.
- Approche recommandée pour les essais de supériorité, car les déviations au protocole ou les changements de traitement vont, en principe, **survenir de façon semblable dans les deux groupes et agir dans le sens d'une homogénéisation des réponses**.
- Dans un essai de supériorité, cette approche est donc **conservatrice** (elle n'augmente pas le risque α).

Population d'analyse ITT (2)

- Dans les essais d'équivalence, cette homogénéisation produit un **effet inverse** en **augmentant** le risque α .
- Une simple approche en ITT est à **éviter** lors des essais d'équivalence.
- Cependant intérêt de permettre la comparaison des essais entre eux (reproductibilité) et de vérifier l'efficacité du traitement de référence.

«Once randomised, always analysed»

Population d'analyse (2) Per Protocole

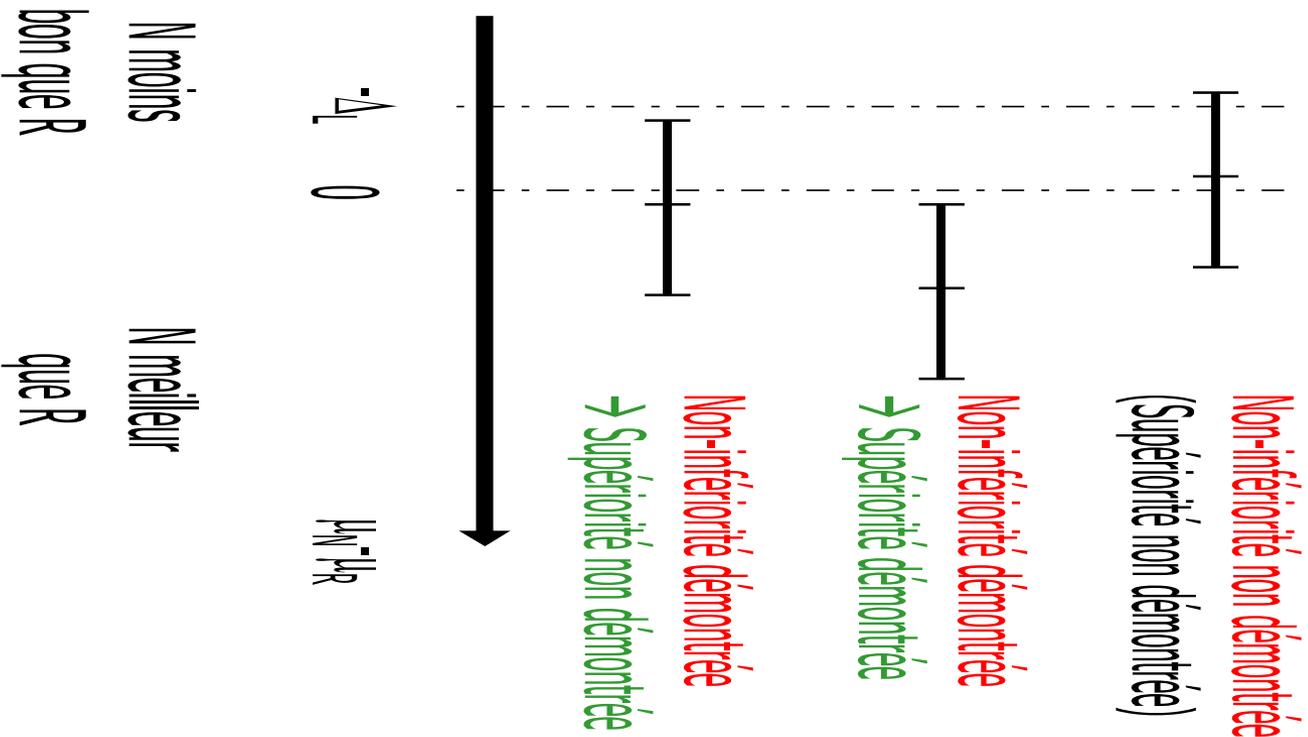
- Analyse per protocole (PP), seuls les patients qui n'ont présenté ni changement ou arrêt de traitement, ni infraction majeure au protocole, sont conservés pour l'analyse.
- On peut s'attendre à une augmentation du contraste entre les groupes pour la réponse au traitement, et donc à une approche conservatrice dans un essai d'équivalence
- **Attention !**
 - Si le traitement expérimental est moins efficace et que l'on retire les patients répondant le plus mal, il ne seront pas analysés en PP, on surestime l'effet du nouveau traitement et l'essai peut à tort conclure à la non infériorité
 - Idem si on retire de l'analyse les patients ayant des effets secondaires importants avec le traitement

Population d'analyse (2) Per Protocole suite

- Donc attention toute particulière à la description des effets secondaires et des écarts au protocole, qui seule permet de suspecter ce type de biais de sélection.
- **Attention** : bien décrire les raisons du choix de cette population, c'est-à-dire les problèmes d'adhérence, de retraits d'étude, de changements de traitement, les toxicités et les déviations majeures au protocole qui ont été à la base du changement de groupe ou de la sortie d'étude de certains sujets.
- Plus la population PP est réduite et déséquilibrée entre les groupes de traitement, plus difficile sera la possibilité de conclure à l'équivalence/non infériorité (manque de puissance).

Passage de la non infériorité à la supériorité

- Les différents cas possibles



Passage de la non infériorité à la supériorité

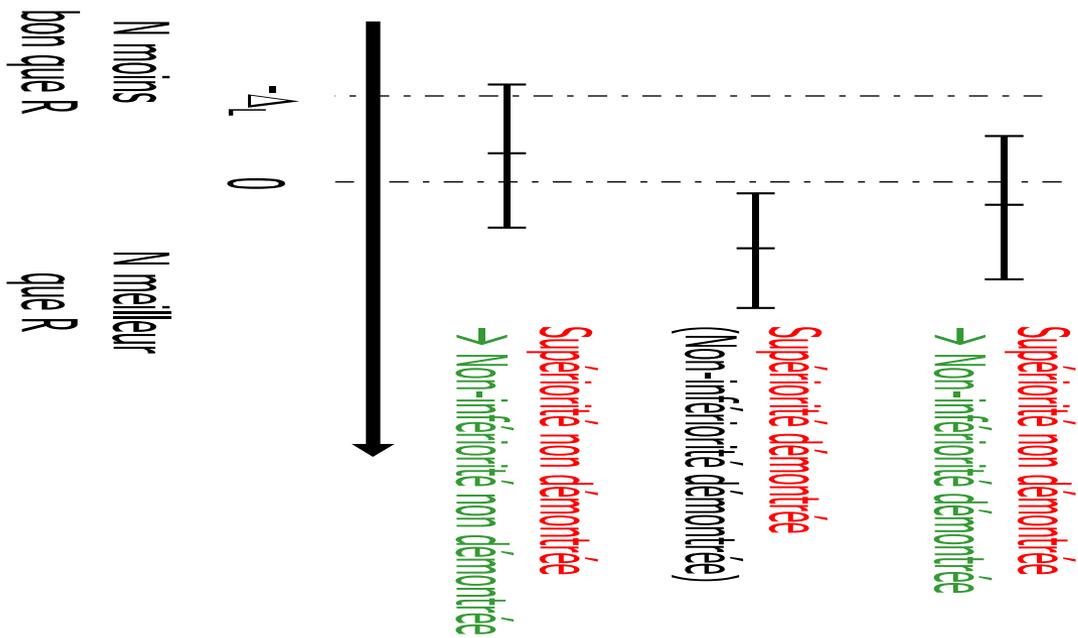
Les conditions pour conclure

- La puissance du test de supériorité doit être suffisante. Les essais de non infériorité nécessitant le plus souvent un nombre de sujets supérieur aux essais de supériorité, ce point est en général vérifié
- Le risque est donc plutôt que le test soit trop puissant et aboutisse à déclarer supérieur le nouveau traitement malgré une différence très petite (et cliniquement non pertinente). Il est donc préférable d'avoir défini *a priori* dans le protocole un delta minimal d'efficacité.
- Le bénéfice clinique supplémentaire doit être évalué en tenant compte du rapport bénéfice/risque
- La population d'analyse privilégiée doit être la population en ITT
- Les tests ou intervalles de confiance doivent être adaptés, en utilisant le même risque α que l'objectif de non infériorité
- Les degrés de signification des tests doivent être fournis pour vérifier la robustesse des résultats

Passage de l'essai de supériorité à la non infériorité

Les conditions pour conclure (1)

- Les différentes situations



Passage de la supériorité à la non infériorité

Les conditions pour conclure (2)

- Le passage d'une conclusion de non supériorité à la non infériorité n'est envisageable que si la non significativité du test est due à une différence observée suffisamment faible et non pas à un manque de puissance.
- Donc si essai de supériorité qui s'est révélé non significatif, il faut calculer *a posteriori* la puissance dont l'essai disposait pour mettre en évidence une certaine différence, que l'on suppose cliniquement insignifiante.
- Si la puissance est assez grande, on peut vouloir conclure à la non infériorité ; par ex : si la puissance *a posteriori* est par exemple de 95% ($\beta=5\%$), et si l'on bascule ensuite dans une démarche de non infériorité, alors ce 5% peut s'interpréter à peu près comme un risque α .

Mais... non satisfaisant

- Essai pensé comme supériorité => risque de biais rendant les groupes *faussement semblables*
- La décision sur le caractère cliniquement négligeable de la différence entre traitements n'est pas le fruit d'une réflexion *a priori*. Elle est portée *a posteriori*, et peut être affectée par la connaissance que l'on a alors des données et de la forme de la relation entre la différence d'effet et la puissance
- **Pas de véritable contrôle du risque α .** Fondamentalement, après avoir eu une réponse décevante à la question initialement posée (est-ce que le traitement N est supérieur au traitement R ?), les expérimentateurs décident *a posteriori*, pour la même étude, de changer la question (est-ce que N et R sont équivalents ?). Démarche à proscrire, car elle augmente de façon non maîtrisable le risque α .

Conclusion

- Recours à ce type d'essai + fréquent
- Traitements efficaces mais besoin d'alternatives
 - Coût
 - Effets secondaires
 - Facilité d'administration
- Non infériorité/équivalence **n'est pas** le miroir des essais de supériorité

JAMA, December 26, 2012—Vol 308, No. 24

**Reporting of Noninferiority
and Equivalence Randomized Trials**
Extension of the CONSORT 2010 Statement

Conclusion (2)

<https://www.fda.gov/media/78504/download>

https://www.ema.europa.eu/en/documents/scientific-guideline/guideline-choice-non-inferiority-margin_en.pdf

Une vidéo du NEJM sur le sujet

<https://www.youtube.com/watch?app=desktop&v=bv9GXUbqyfs>



THINGS GOT REALLY INTERESTING
WHEN THE STATISTICIAN STARTED
DOING WARD ROUNDS

Un exemple, comparaison de proportion

True proportion in active control	Non-inferiority bound using 10% margin	Approximate sample size per group assuming 1:1 randomization to new treatment and control required under:		
		Equal effects	5% benefit	10% benefit
0.1	0.09	19,200	8,725	5,050
0.2	0.18	8,500	3,900	2,250
0.3	0.27	4,970	2,260	1,300
0.4	0.36	3,200	1,450	825
0.5	0.45	2,100	1,000	550
0.6	0.54	1,440	640	360
0.7	0.63	930	405	225

Sensibilité d'analyse

- Faculté de l'essai à distinguer un traitement efficace d'un traitement inefficace
- Dépend de la taille d'effet que l'on désire détecter
- Non mesuré dans un essai de non infériorité
- Difficultés
 - Mauvaise observance
 - Amélioration spontanée
 - Traitements concomitants
 - Critères diagnostiques