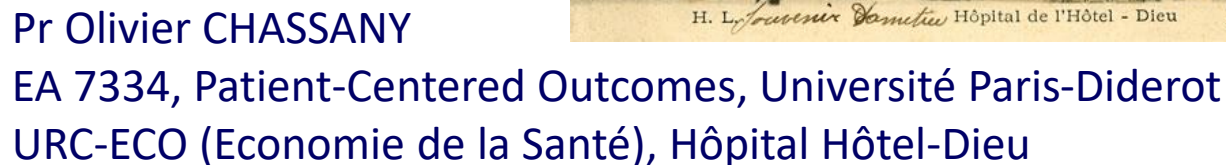


ASSISTANCE  HÔPITAUX  
PUBLIQUE DE PARIS



# Comparabilité des groupes à l'inclusion

COMPARABILITÉ DES GROUPES À L'INCLUSION sur des données démographiques, diagnostiques, pronostiques

- Faut-il faire des comparaisons statistiques qui augmentent le risque alpha (démontrer à tort une différence qui n'existe pas) ?
- Inversement une différence peut être pertinente mais non significative car peu de patients sont inclus (risque bêta = manque de puissance)
  - Quand le poids à l'inclusion est de  $90 \pm 5$  kg dans le groupe traité et de  $101 \pm 4$  kg dans le groupe placebo, mais p non significatif...

Noble, RE. A six-month study of the effects of dexfenfluramine on partially successful dieters. Curr Ther Res. 1990;47:612–619

# Analyse en intention de traiter (Intent to treat analysis - ITT)

## Principe de l'analyse en ITT ("full set analysis") :

- Prise en compte de tous les patients randomisés (pour garder la comparabilité des groupes randomisés)
- Même ceux présentant des déviations au protocole :
  - Pas de prise correcte du traitement
  - Arrêt prématuré du traitement
  - Sortie d'essai, perdus de vue...

Cela nécessite de remplacer les données manquantes

# Analyse en intention de traiter (Intent to treat analysis – ITT)

**Il est licite de ne pas prendre en compte ("modified full set analysis") :**

- Patient inclus à tort
- Patient n'ayant pris aucune dose du traitement
- Patient pour lequel il n'existe aucune donnée d'évaluation après la randomisation

# Analyse en intention de traiter (Intent to treat analysis - ITT)

- Le remplacement des données manquantes est toujours délicat :
  - Techniques plus ou moins complexes
    - Depuis la simple dernière évaluation disponible (LOCF "last observation carried forward")
    - La valeur lors de l'inclusion
    - Échec ou décès
    - Jusqu'à des techniques d'imputation et de modélisation
    - (*Extrême : hypothèse du biais maximum*)
- Le meilleur moyen de résoudre la question des données manquantes est de ne pas en avoir...

# Analyse en intention de traiter (Intent to treat analysis - ITT)

- Robuste : tendance à diminuer la différence entre les groupes de traitement
- Si différence statistiquement significative : confiance dans les résultats
- **Pourquoi faire une analyse ITT, qui inclut des patients qui n'ont probablement pas bien répondu au traitement ?**
  - Car exclure de l'analyse des patients pour différentes raisons, risque de détruire la comparabilité des groupes
  - En ne gardant que les "bons" patients (analyse Per Protocol), on peut favoriser le traitement à l'étude vs traitement comparateur

# Analyse per protocol (PP)

## Avantages

- Ne prend en compte que les patients ayant correctement pris le traitement jusqu'à son terme, sans déviation au protocole
- Plus précis pour démontrer la vraie efficacité du traitement

# Analyse per protocol (PP)

## Inconvénients

- Risque d'introduire un biais dans la comparabilité des groupes
  - Doute possible sur la validité de l'essai
  - **Analyse moins robuste, ce d'autant que le nombre de patients exclus de l'analyse est grand**
- À l'extrême, en excluant les patients n'ayant probablement pas répondu au (nouveau) traitement :
  - On augmente la différence entre les 2 traitements
  - On maximalise la probabilité de montrer une différence statistiquement significative (essai de supériorité)

# Analyse en intention de traiter ou per protocol (PP) ?

|                                                                                                                                                                                                                                                                                                            | Supériorité                                                   | Non-infériorité                                         |
|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---------------------------------------------------------------|---------------------------------------------------------|
| <b>Analyse en intention de traiter (ITT)</b> <ul style="list-style-type: none"><li>Tendance à réduire la différence observée (surtout s'il y a eu beaucoup de déviations au protocole, sorties d'essais dans les 2 groupes...)</li></ul>                                                                   | <b>Méthode la plus robuste pour démontrer une supériorité</b> | Risque de conclure à tort à l'équivalence               |
| <b>Analyse "per protocol" (PP)</b> <ul style="list-style-type: none"><li>Censée optimiser une différence entre 2 traitements, si celle-ci existe</li><li>Mais, le nombre de patients exclus de l'analyse ne doit pas être trop important, ni déséquilibré entre les 2 groupes (nombre et motifs)</li></ul> | Seulement pour confirmer les résultats en ITT                 | Possiblement plus adaptée à un essai de non-infériorité |

# Analyse en intention de traiter ou per protocol (PP) ?

|                                       | Supériorité                                     | Non-infériorité                                                                                  |
|---------------------------------------|-------------------------------------------------|--------------------------------------------------------------------------------------------------|
| Analyse en intention de traiter (ITT) | Seule analyse non critiquable pour l'AMM<br>+++ | Présenter les résultats en IT et PP pour vérifier leur cohérence en faveur de la non-infériorité |
| Analyse "per protocol" (PP)           |                                                 |                                                                                                  |

# Multiplicité des tests statistiques

**Augmente le risque alpha de montrer à tort une différence qui n'existe pas**

- Critère principal mesuré plusieurs fois au cours du temps :
  - Fixer le temps principal (6 mois, 1 an...)
- Multiples critères (principal, secondaires, tertiaires)
  - Définir le critère principal, le premier critère secondaire
- Hiérarchie des critères d'évaluation

# Overall Risk of Type I Error in Using Repeated $t$ Tests at $\alpha = 0.05$

**Table 11.1.2** Overall risk of Type I error in using repeated  $t$  tests at  $\alpha = 0.05$

| $I$ | Overall risk |
|-----|--------------|
| 2   | 0.05         |
| 3   | 0.12         |
| 4   | 0.20         |
| 6   | 0.37         |
| 8   | 0.51         |
| 10  | 0.63         |

# Risque de reconnaître à tort un bénéfice sur des critères d'évaluation secondaires

## **Multiplicité des tests par multiplication des critères d'évaluation secondaires**

- Tentative de rattraper un résultat non significatif sur le critère principal
- Tendance à revendiquer des effets additionnels au critère principal déjà significatif (identification de facteurs de différenciation)

# Un seul critère d'évaluation ?

## ➔ critère d'évaluation UNIQUE (principal)

Aucun bénéfice reconnu en dehors du critère primaire  
On ne peut plus tricher car protocoles enregistrés dans  
clinicaltrials.gov



REDUCTEUR de résumer en un seul critère le  
bénéfice d'un traitement

Exemple : OS et PFS en cancérologie

# Solution : co-critères primaires

## Plusieurs critères et partage du risque alpha sur ces CJ

Ajustement du seuil (Méthodes classiques : Hochberg, Bonferroni)

**Ex : 2 co-critères primaires et risque d'erreur global de 5%**

- Critère 1 : seuil  $\alpha$  0.025                      seuil  $\alpha$  0.046 OS
- Critère 2 : seuil  $\alpha$  0.025                      seuil  $\alpha$  0.004 PFS

Gilbert MR. N Engl J Med 2014



**conclure indépendamment sur chacun des critères**



**nombre limité de co-critères car ajustement pénalisant**

➔ Abaissement du risque  $\alpha$  attribué à chaque critère

➔ Augmentation du nombre de patients

## Autre solution : méthode séquentielle hiérarchique (gatekeeper)

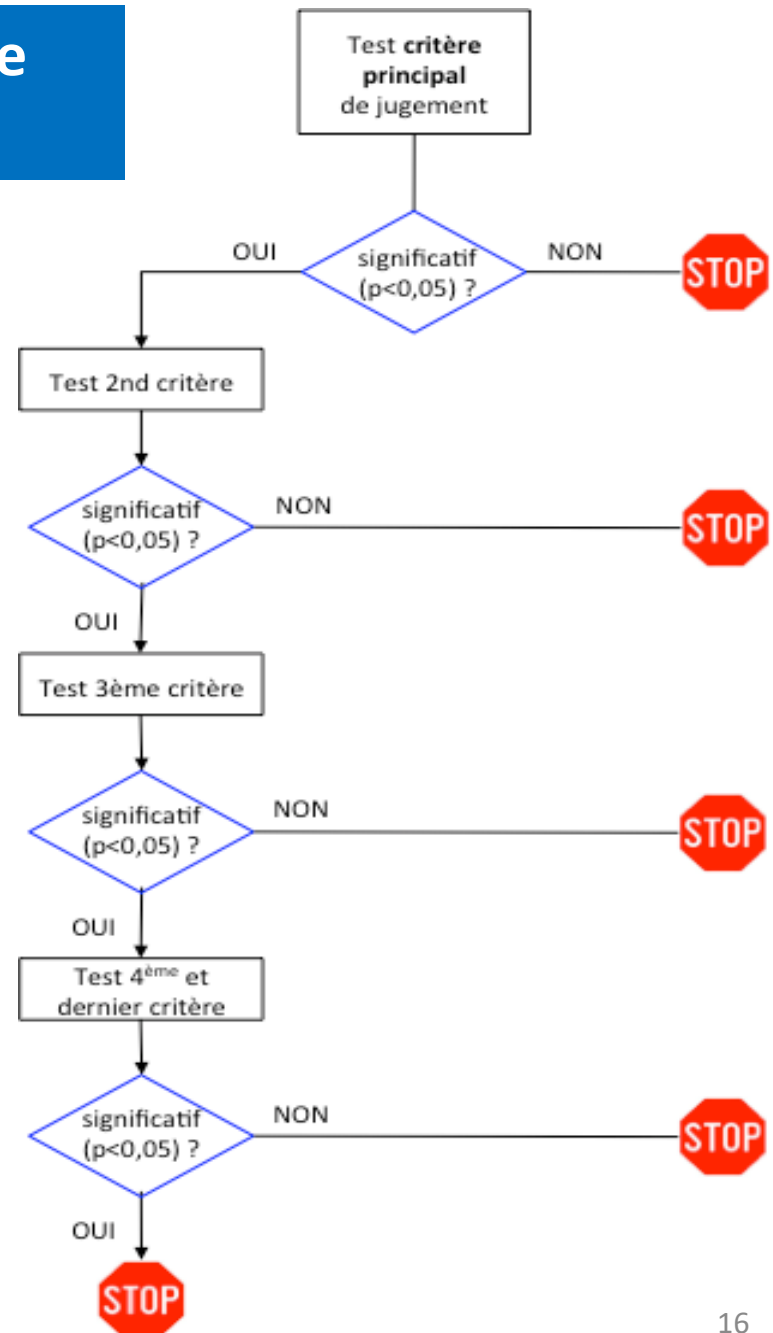
1. Hiérarchisation a priori
2. Tests séquentiels
3. Fin de la procédure au premier test non significatif

### Intérêt

**Du fait de l'emboîtement :**

**Pas d'inflation du risque  $\alpha$**

→ Démonstration licite des avantages supplémentaires



# Pas de comparaison avant-après dans un essai randomisé

**PAS DE COMPARAISON AVANT-APRÈS** dans chaque groupe sans comparaison entre les 2 groupes :

- *Exemple : l'évolution entre le début et la fin du traitement est statistiquement significative dans le groupe traité, alors que dans le groupe placebo, il n'y a pas de différence...*

Mais on prend habituellement la valeur de baseline en co-variable :

- Comparaison  $\Delta_A$  (avant-après) versus  $\Delta_B$  (avant-après)

# Analyse en sous-groupe

Dogme : Seules pourront être prises en compte des analyses en sous-groupe :

1. Planifiées à priori
2. Et si il existe une différence significative sur le critère principal sur l'ensemble des patients

**TABLEAU II - RÉPONSE AU TRAITEMENT PAR NÉBULISATION  
CHEZ LES ASTHMATIQUES ET LES BPCO**

| -                                             | Salbutamol | Salbutamol<br>+ BI | Valeur de $P^{\dagger}$<br>(IC à 95%) $^{\ddagger}$ |
|-----------------------------------------------|------------|--------------------|-----------------------------------------------------|
| <i>Patients au total</i>                      |            |                    |                                                     |
| Effectif                                      | 44         | 59                 |                                                     |
| DEP moyen de base (l/min)                     | 128        | 128                |                                                     |
| DEP moyen à 1 h (l/min)                       | 163        | 184                | 0,12                                                |
| Augmentation moyenne<br>du DEP (%)            | 28         | 52                 | (1-48%)                                             |
| <i>Total asthmatiques</i>                     |            |                    |                                                     |
| Effectif                                      | 23         | 33                 |                                                     |
| DEP moyen de base (l/min)                     | 145        | 137                |                                                     |
| DEP moyen à 1 h (l/min)                       | 190        | 222                | 0,027                                               |
| Augmentation moyenne<br>du DEP (%)            | 31         | 77                 | (8-84%)                                             |
| <i>Asthmatiques à DEP<br/>&lt;140 l/min *</i> |            |                    |                                                     |
| Effectif                                      | 12         | 23                 |                                                     |
| DEP moyen de base (l/min)                     | 98         | 96                 |                                                     |
| DEP moyen à 1 h (l/min)                       | 129        | 179                | 0,04                                                |
| Augmentation moyenne<br>du DEP (%)            | 31         | 93                 | (4-121%)                                            |
| <i>BPCO</i>                                   |            |                    |                                                     |
| Effectif                                      | 21         | 26                 |                                                     |
| DEP moyen de base (l/min)                     | 109        | 116                | 0,55                                                |
| DEP moyen à 1 h (l/min)                       | 134        | 135                |                                                     |

**= 103, mais 125  
patients  
randomisés, où  
sont 22 patients ?**

\* À l'admission.

$^{\dagger}$  Test U de Mann Whitney, 1 h contre valeur de base ;  $P > 0,05$  = non significatif.

$^{\ddagger}$  Pour la différence.

# Aérosol de salbutamol avec ou sans bromure d'ipratropium dans l'obstruction aiguë des voies aériennes

|                                               |                                                                                                                                                                                                                                                                                      |
|-----------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Différence significative : <b>NON</b>         | Il n'y a pas de différence statistique entre les 2 groupes (44 vs 59 patients: 184 vs 163 de peak flow, $p = 0,12$ )                                                                                                                                                                 |
| Analyse en sous-groupe : <b>PAS DE VALEUR</b> | A partir du moment où la comparaison globale est non significative, et que les sous-groupes n'ont pas été constitués au départ ( <b><u>stratification</u></b> ). La répartition des traitements dans les groupes « asthme » et « bronchite chronique » est d'ailleurs déséquilibrée. |

O'Driscoll BR, et al. Nebulised salbutamol with and without ipratropium bromide in acute airflow obstruction. Lancet. 1989 Jun 24;1(8652):1418-20.

## ET SI LA DIFFÉRENCE N'EST PAS SIGNIFICATIVE : manque de puissance ? ( $\beta$ trop élevé)

- $H_0$  nulle : hypothèse d'égalité des traitements
- **$H_1$  alternative : traitement A supérieur au traitement B**

|          |                                | REALITE  |                                |
|----------|--------------------------------|----------|--------------------------------|
|          |                                | HO       | H1                             |
| DECISION | NON REJET<br>HO ( $p > 0.05$ ) |          | $\beta$ manque de<br>Puissance |
|          | REJET HO<br>(H1) $p < 0.05$    | $\alpha$ | 1- $\beta$ puissance           |

Risque  $\alpha$  : montrer à tort une différence ( $\Delta$ ) qui n'existe pas

Risque  $\beta$  : ne pas pouvoir montrer une  $\Delta$  qui existe réellement

## **Et si la différence est significative**

- **Biais (validité interne) ?**
- **Pertinence de la différence ?**
- **Validité externe (extrapolation)**

# Biais

| Source de biais                                                      | Solution                                                            |
|----------------------------------------------------------------------|---------------------------------------------------------------------|
| • <b>Confusion</b>                                                   | Comparaison groupe contrôle (respectée ?)                           |
| • <b>Sélection</b>                                                   | Randomisation (respectée ?)                                         |
| • <b>Suivi et d'évaluation</b><br><i>(Performance and detection)</i> | Double-aveugle (respecté ?)                                         |
| • <b>Attrition (perte de comparabilité)</b>                          | Suivi rigoureux des patients ?<br>Analyse en intention de traiter ? |
| Étude de cohorte                                                     |                                                                     |
| • <b>Confusion</b>                                                   | Analyse multivariée                                                 |

## Pertinence de la différence ?

- 2% de réduction de la mortalité
- 3 semaines de survie
- 0,5% d'hémorragies graves en moins

# Treatment of symptoms of moderate to severe irritable-bowel syndrome (IBS) with constipation

| Linacotide vs placebo                                                                                    | 12-wk A                      | 12-wk B                       | 26-wk B                      |
|----------------------------------------------------------------------------------------------------------|------------------------------|-------------------------------|------------------------------|
| Abdominal pain responder (30%)                                                                           | 55 vs 42%<br>$\Delta = 13\%$ | 54 vs 39%<br>$\Delta = 15\%$  | 54 vs 36%<br>$\Delta = 18\%$ |
| IBS relief responder                                                                                     | 37 vs 19%<br>$\Delta = 16\%$ | 39 vs 17%<br>$\Delta = 16\%$  | 37 vs 17%<br>$\Delta = 20\%$ |
| Abd pain responder (40%)<br>Abd pain responder (50%)                                                     | 48 vs 34%<br>39 vs 26%       | <b>← Sensitivity analyses</b> |                              |
| Sustained <ul style="list-style-type: none"> <li>Pain responder</li> <li>IBS relief responder</li> </ul> | 53 vs 41%<br>34 vs 18%       |                               |                              |

- Especially for the EU analysis, analysis of sustained response : overall 50% response with an additional 50% response during the last 4 weeks thus excluding a deterioration towards the end of treatment

**Étude de prévention secondaire chez 4444 patients (angine de poitrine, infarctus secondaire) traités pendant 5,4 ans par simvastatine ou placebo**

| <b>Taux de mortalité à 5 ans</b>                                   | <b>Placebo<br/>12,4%</b>                              | <b>Simvastatine<br/>8,7%</b> | <b>Calcul</b>      |
|--------------------------------------------------------------------|-------------------------------------------------------|------------------------------|--------------------|
| <b>Diminution absolue du taux de mortalité à 5 ans</b>             | <b><math>\Delta = 3,7\%</math></b>                    |                              | <b>(12,4-8,7)</b>  |
| Diminution relative du taux de mortalité à 5 ans                   | 30%                                                   |                              | (3,7/12,4)         |
| Risque relatif (Relative Risk)                                     | RR = 0,70<br>IC <sub>95</sub> = 0,58-0,83, p = 0,0003 |                              | (8,7/12,4)         |
| <b>Nombre de patients à traiter pendant 5 ans pour éviter un †</b> | <b>27</b>                                             |                              | <b>(100 / 3,7)</b> |

Randomised trial of cholesterol lowering in 4444 patients with coronary heart disease : the Scandinavian simvastatin survival study (4S). Lancet 1994.

# Extrapolation des résultats

Fonction de :

- Critères d'inclusion et de non inclusion
- Comparaison des résultats à d'autres études