

UE Santé Publique FGSM3

LCA des études interventionnelles

Giuseppe Balice, MD, MSc, PhD(cand)
11 septembre 2026



Table des matières

1. Inclusion et randomisation
2. Biais de sélection et biais de mesure
3. Objectif de la comparaison : supériorité et non-infériorité
4. Les critères de jugement : durs, multiples, intermédiaires
5. Le risque d'erreur de premier type

Code couleur du texte de ce diaporama

1. Inclusion et randomisation
2. Biais de sélection et biais de mesure Biais (systématique), effet (aléatoire)
3. Objectif de la comparaison Supériorité
4. Les critères de jugement Une seule analyse
5. Le risque d'erreur de premier type Analyses multiples
6. Tout le diaporama Emphase

Définition de la question de recherche

P opulation

F actor (Intervention)

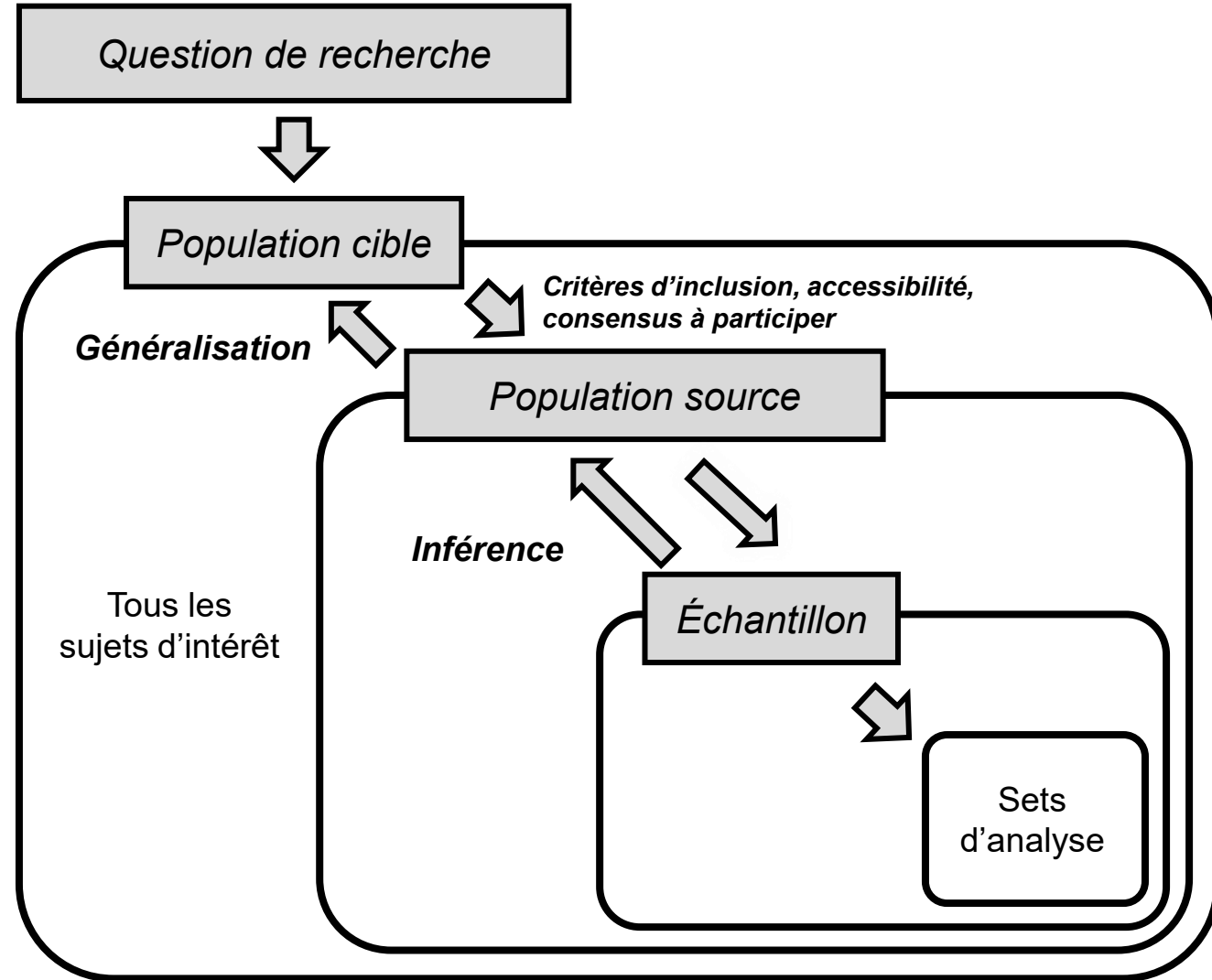
C omparator

O utcome

- Les mots « question de recherche », « hypothèse » et « objectif » décrivent la même chose, **déclinée linguistiquement** en tant que question, doute ou affirmation
- Le critère de jugement est donc **un moyen pour atteindre l'objectif** ; en particulier, ce qu'on mesure chez les sujets inclus dans l'étude

La démarche hypothético-déductive

1. Formuler la **question** de recherche (PICO)
2. Réfléchir à la meilleure **méthode** pour répondre à la question (rédiger le protocole de recherche)
3. Mettre en **pratique** le protocole (inclusion, investigation, analyse, publication des résultats)
4. Répondre à la question de recherche (**lecture critique** des résultats)

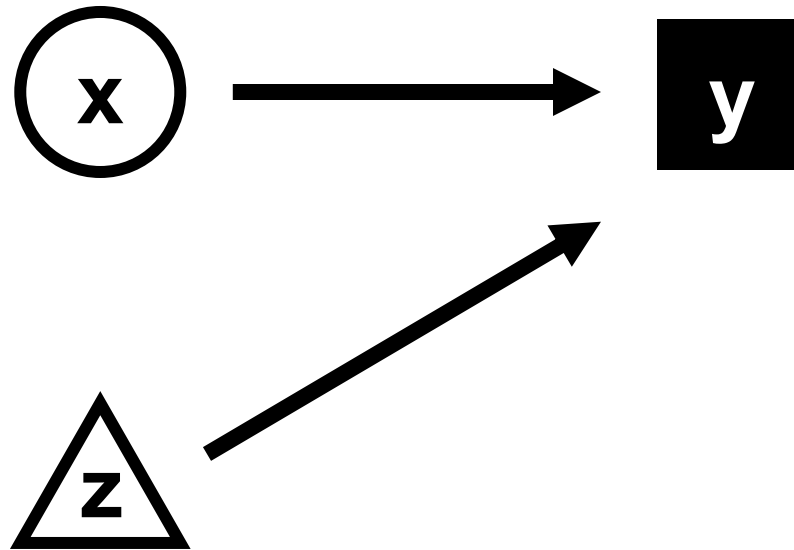


Hypothèse de recherche : questions type

- 1. Quelles sont les données préliminaires mentionnées par les auteurs ? (L'hypothèse de recherche tient la route?)**
- 2. Quel est l'objectif principal de l'étude ?**

Essai interventionnel

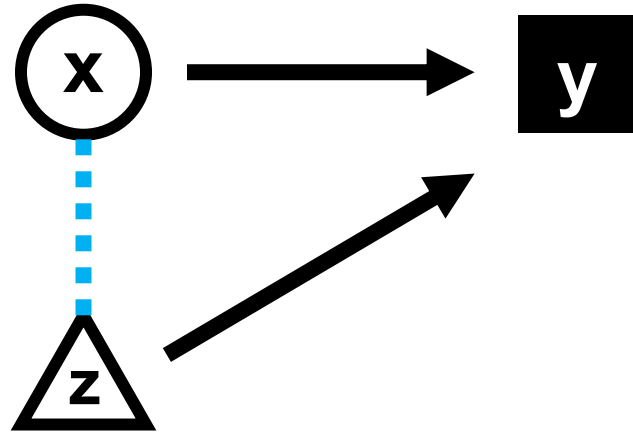
L'objectif de l'essai interventionnel est d'estimer **l'effet du traitement** administré (x) sur un critère de jugement (y)...



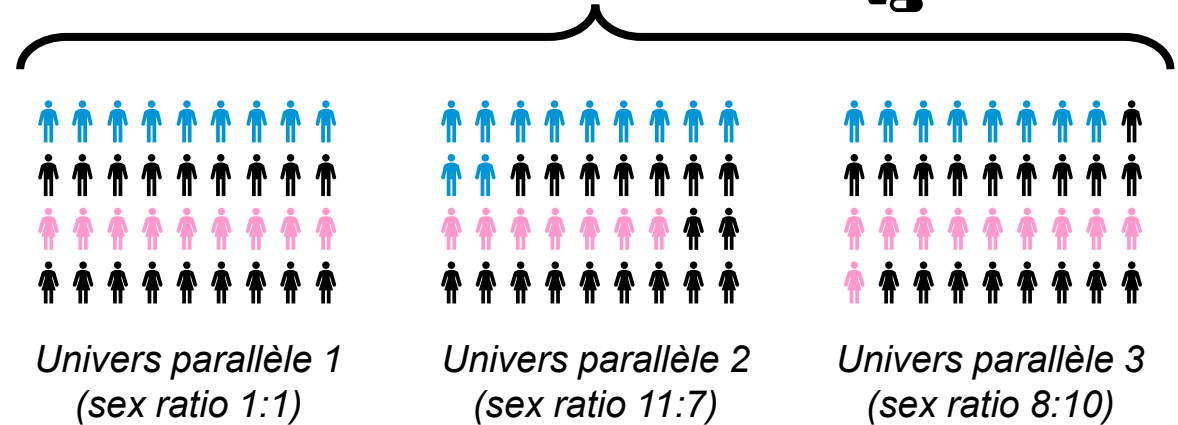
...mais (y) est influencé par plein d'autres facteurs (par exemple, z) !

Explorons tous les cas possibles dans les univers parallèles...

Si z n'influence pas le choix d'administrer le traitement à mon patient :

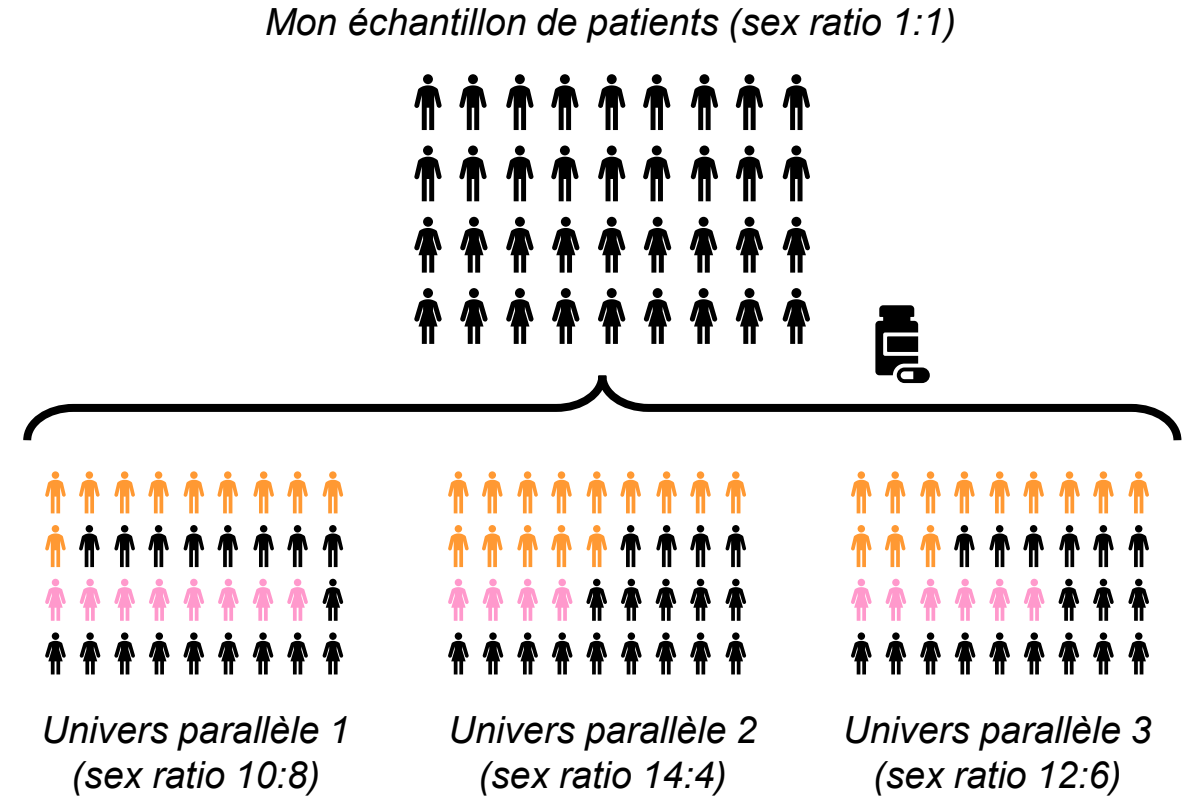
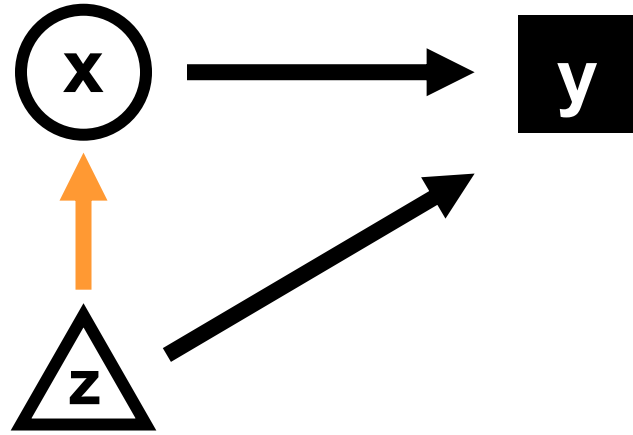


Mon échantillon de patients (sex ratio 1:1)



- Il peut être plus **ou moins** fréquent dans le groupe intervention, **par hasard**.
- Dans les univers parallèles 2 et 3, le Δy mesuré entre les groupes n'est pas dû seulement à x , mais aussi au déséquilibre de z (**effet de confusion**, mais ce n'est pas un **biais** de confusion, car il n'est pas **systématique** dans tous les univers parallèles). Donc, on mesure l'effet $x \rightarrow y$ avec une **erreur aléatoire**.
- Pour corriger cette erreur, on peut utiliser des méthodes statistiques.

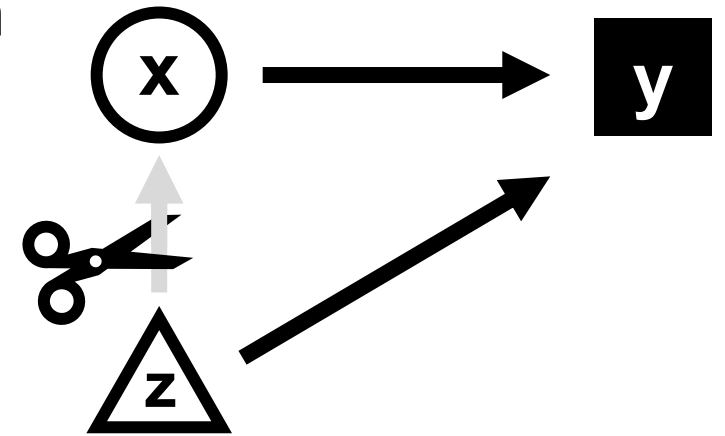
Si **z influence le choix** d'administrer le traitement à mon patient :



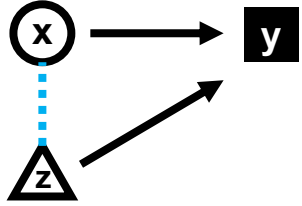
- Il sera **systematiquement** plus fréquent dans le groupe intervention.
- Dans tous les univers parallèles il sera **impossible de démêler** la contribution de **x** de celle de **z** au Δy mesuré entre les groupes. Cela entraine un **biais de sélection**.
- Définition scolastique du biais de sélection de l'étude interventionnelle : le groupe intervention est systématiquement différent du groupe contrôle pour un certain facteur.

Comment prévenir le biais de sélection ?

- **Prérequis éthique** : **clause d'ambivalence** (equipoise). On ne sait pas si le traitement est plus ou moins efficace que le comparateur.
- **Solution** : **la randomisation**. Le fait de recevoir un traitement dépend seulement du hasard.
- **Résultat** : **par construction**, nous savons que la différence entre les effets $x \rightarrow y$ mesurés dans les univers parallèles est due **seulement** à une erreur aléatoire, et non pas à un biais. Nous pouvons donc la gérer avec des méthodes statistiques.



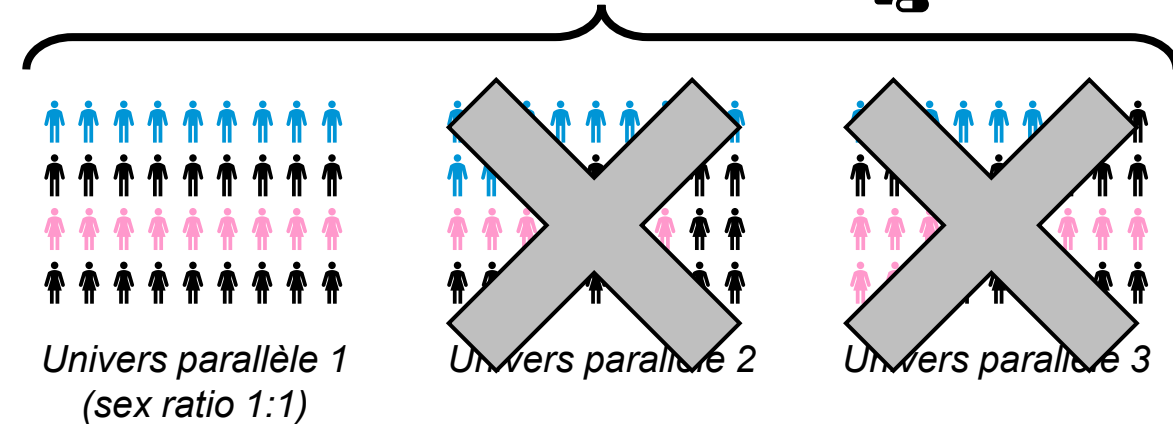
La randomisation stratifiée



- L'effet de confusion peut être contrôlé avec des techniques d'analyse statistique : analyse multivariée, analyse stratifiée

- Cependant, ces techniques sont « **gourmandes** » en information : il faut beaucoup de patients pour qu'elles estiment correctement l'effet du traitement, en le nettoyant de l'effet de confusion
- Nous pouvons leur faciliter la tâche (↑ **précision de l'estimation**) en randomisant **un nombre strictement égal de patients pour chaque catégorie d'un facteur** de stratification

Mon échantillon de patients (sex ratio 1:1)



La randomisation stratifiée

Si on veut stratifier sur peu de facteurs, on peut utiliser des listes de randomisation préétablies : une liste pour chaque combinaison des facteurs de stratification. Exemples :

- Stratification sur le sexe seul (2 catégories) : 2 listes
- Stratification sur sexe et âge (2 catégories) : 4 listes
- Stratification sur n variables avec k_n catégories : $\prod_n k_n$ listes
- Ça peut monter assez vite !

	1	2	3	4	5	6	7	8	9	10	11	12
H												
F												

Et puisque la randomisation produit des groupes équilibrés seulement avec des grands nombres, nous sommes vite coincés : si je rajoute trop de facteurs de stratification, **le nombre total de patients à inclure explose**.

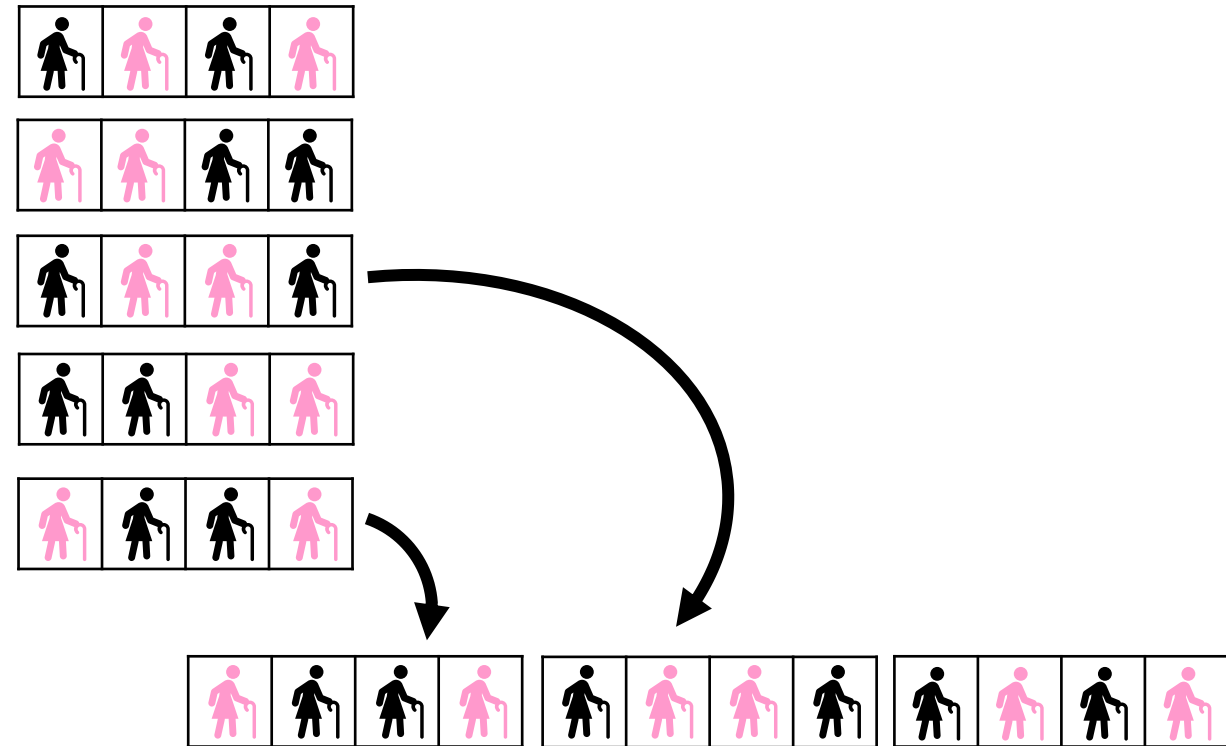
Ce problème est résolu par la minimisation ou par la randomisation en blocs.

		1	2	3	4	5	6	7	8	9	10	11	12
J	H												
	F												
A	H												
	F												

La randomisation par blocs

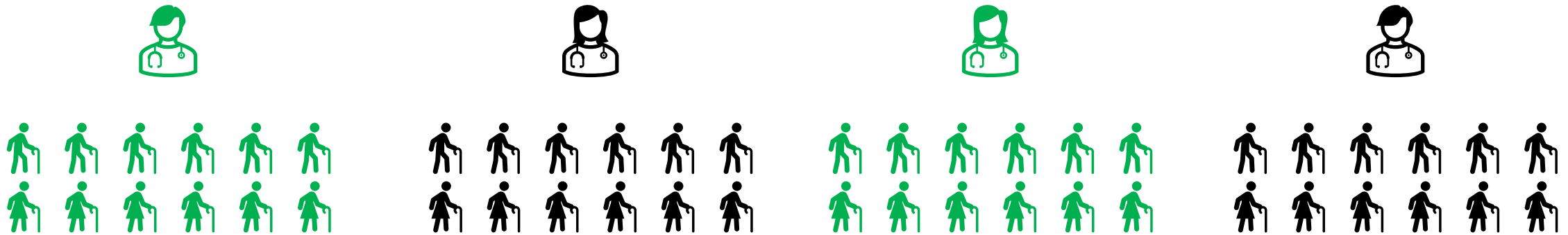
- Dans cet exemple de randomisation précédente, je devais **attendre** de randomiser 12 patients dans chaque liste pour atteindre **l'équilibre** entre les groupes (6 traités et 6 contrôles).
- Je pourrais obtenir le même résultat en mélangeant 3 blocs de taille 4, issus d'un répertoire de **blocs pré équilibrés**
- De plus, si les blocs on une taille variable, le **secret de la randomisation** est renforcé (et donc le risque de biais de sélection est ultérieurement réduit).

		1	2	3	4	5	6	7	8	9	10	11	12
A	F												

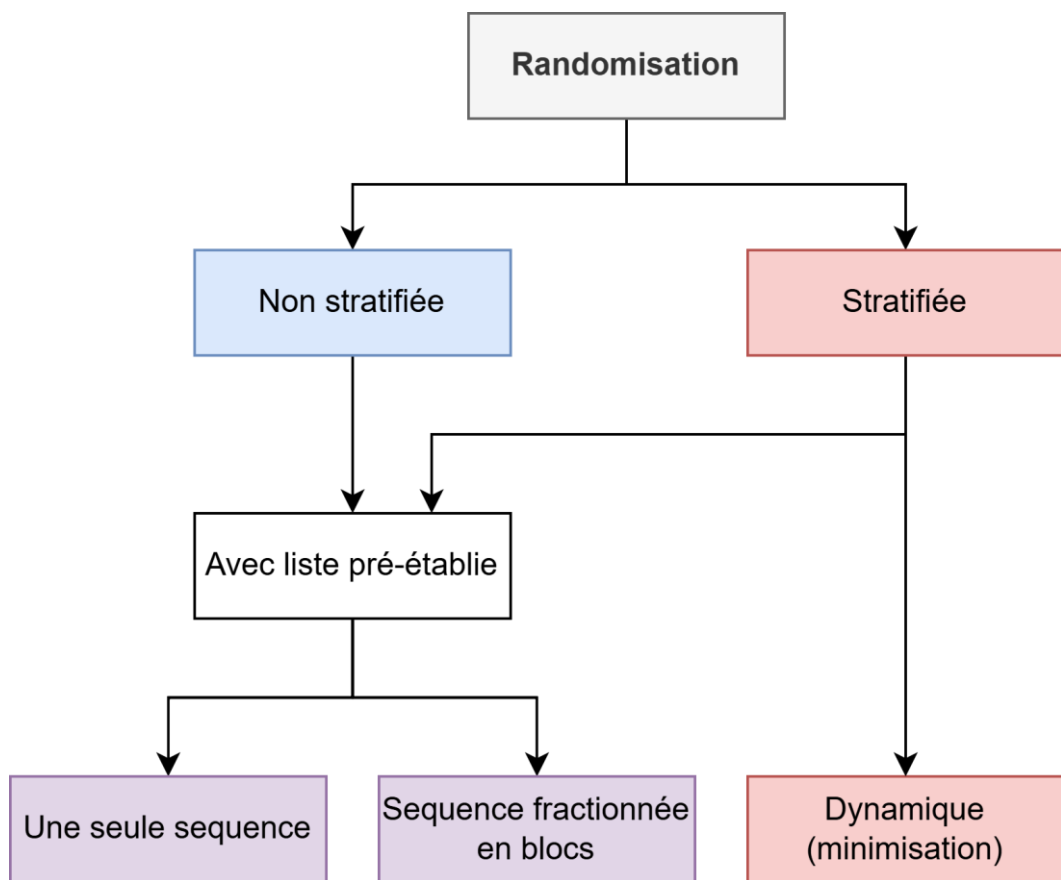


La randomisation par grappes (cluster)

- Dans certains designs expérimentaux (évaluation de pratiques, sans aveugle de l'investigateur) il y a un fort risque de **biais de contamination** : le groupe contrôle pourrait profiter du changement inconscient des habitudes de l'investigateur, induites par sa connaissance du protocole d'intervention
- Pour prévenir ce biais, il faut randomiser en grappe **tous les patients suivis par un seul investigateur**



Le cluster devient l'unité de randomisation, mais le critère de jugement est mesuré toujours au niveau du patient. Il faudra donc **inclure plus de patients** par rapport à un design non en grappes !



Technique de randomisation	Intérêt
Stratifiée	Mettre le modèle statistique dans la meilleure condition pour nettoyer l'effet traitement de l'effet de confusion
Minimisation	Stratifier sur beaucoup de facteurs en même temps, sans faire exploser le nombre de patients à inclure
Par blocs	- Préserver le secret de la randomisation - Garantir l'équilibre des groupes même en cas de faible effectif
Par grappes	Prévenir le biais de contamination

Présentation des patients randomisés : le *Tableau 1*

- Dans un essai contrôlé **randomisé**, il n'y a pas besoin d'afficher une p-value sur chaque ligne du Tableau 1. Le « *eyeball test* » (contrôle au coup d'œil) est suffisant.

Table 1. Characteristics of the Patients at Baseline.*		
Characteristic	Empagliflozin (N = 3304)	Placebo (N = 3305)
Age — yr	63.9±13.9	63.8±13.9
Female sex — no. (%)	1097 (33.2)	1095 (33.1)

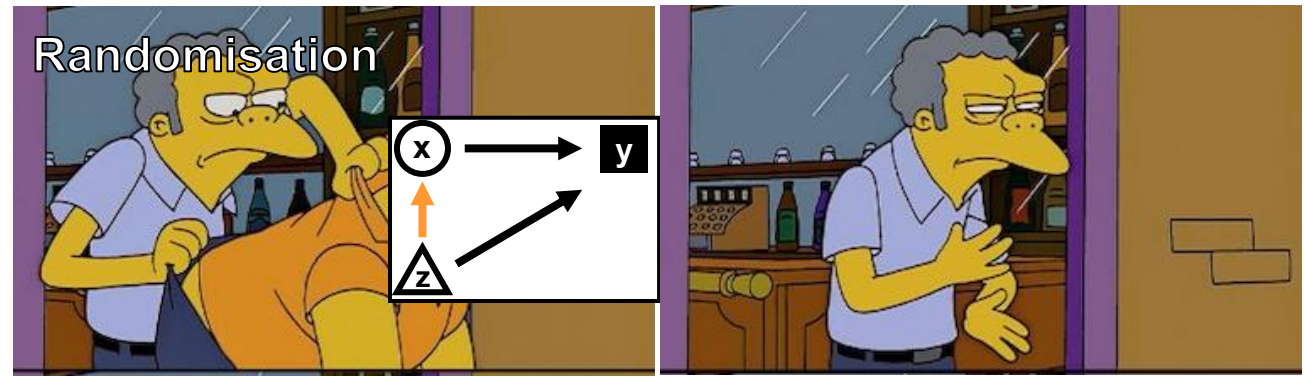
- En effet les deux bras, bien que « comparables **par construction** », ne peuvent pas être strictement identiques pour tous les facteurs mesurés ! (persistance d'une **fluctuation aléatoire**)
- Il est beaucoup plus important de lire le paragraphe des méthodes et critiquer la façon dans laquelle la randomisation a été mise en place.

Attention : cela n'est plus vrai lorsqu'il n'y a pas de randomisation (donc dans toutes les études **observationnelles**). Des méthodes statistiques avancées (score de propension) peuvent être mises en place pour limiter le biais de confusion. Nous en reparlerons prochainement.

Inclusion et randomisation : questions type

- 1. Quels sont les critères d'inclusion de l'essai ?** (La population source est adéquate à la question de recherche ?)
- 2. La randomisation a été bien réalisée ?** (Y-a-t il encore un risque de biais de sélection?)
- 3. Interprétez le Tableau 1.** (Y-a-t il un effet de confusion résiduelle sur une variable importante qu'il faudra prendre en compte dans l'analyse statistique?)

Une fois la randomisation faite, il y a deux problèmes qui peuvent survenir pendant le déroulement de l'essai :

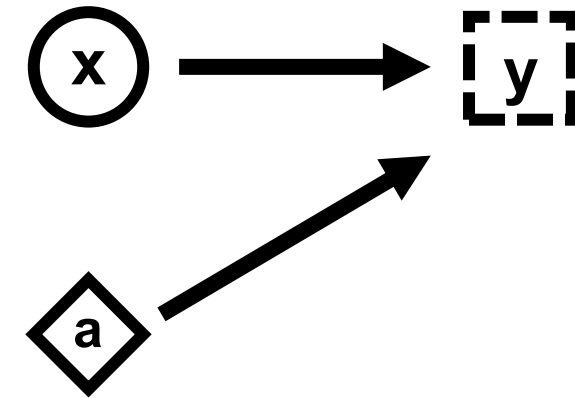


Des patients changent de bras



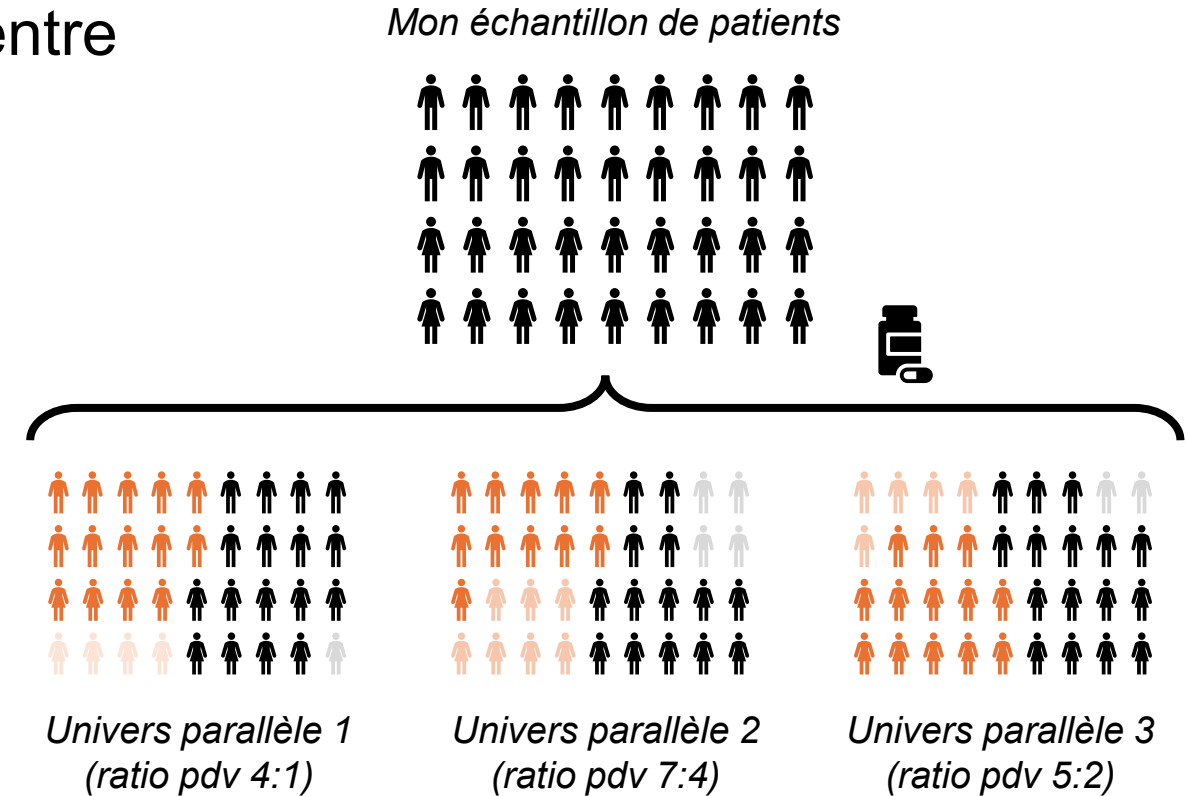
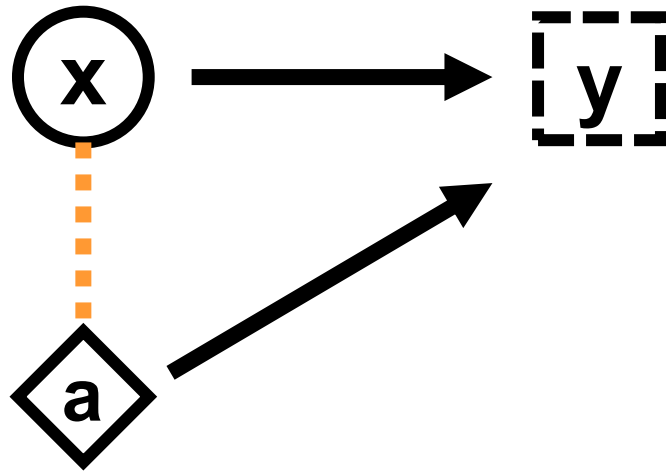
Donc le choix d'avoir changé de bras est influencé par un facteur z (biais de sélection)

Des patients sont perdus de vue



(a) est le fait d'être perdu de vue. Donc on n'observera jamais (y) chez les patients qui présentent le facteur (a).

Devant un **déséquilibre** des perdus de vue entre les deux bras, il faut toujours se demander :



Quel est le risque que le fait d'être perdu de vue ne soit pas simplement lié au hasard, mais qu'il relève d'une différence **systématique** entre les groupes (biais de sélection) ?

Le biais d'attrition est un sous-type de le biais de sélection

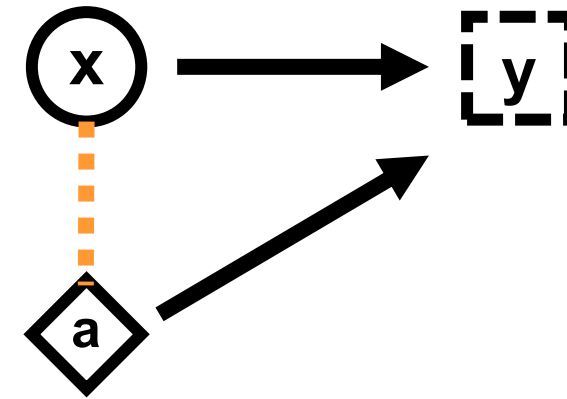
Des patients changent de bras



Analyse en intention de traiter

On respecte l'allocation définie par la randomisation, même si le patient a changé de bras ensuite.

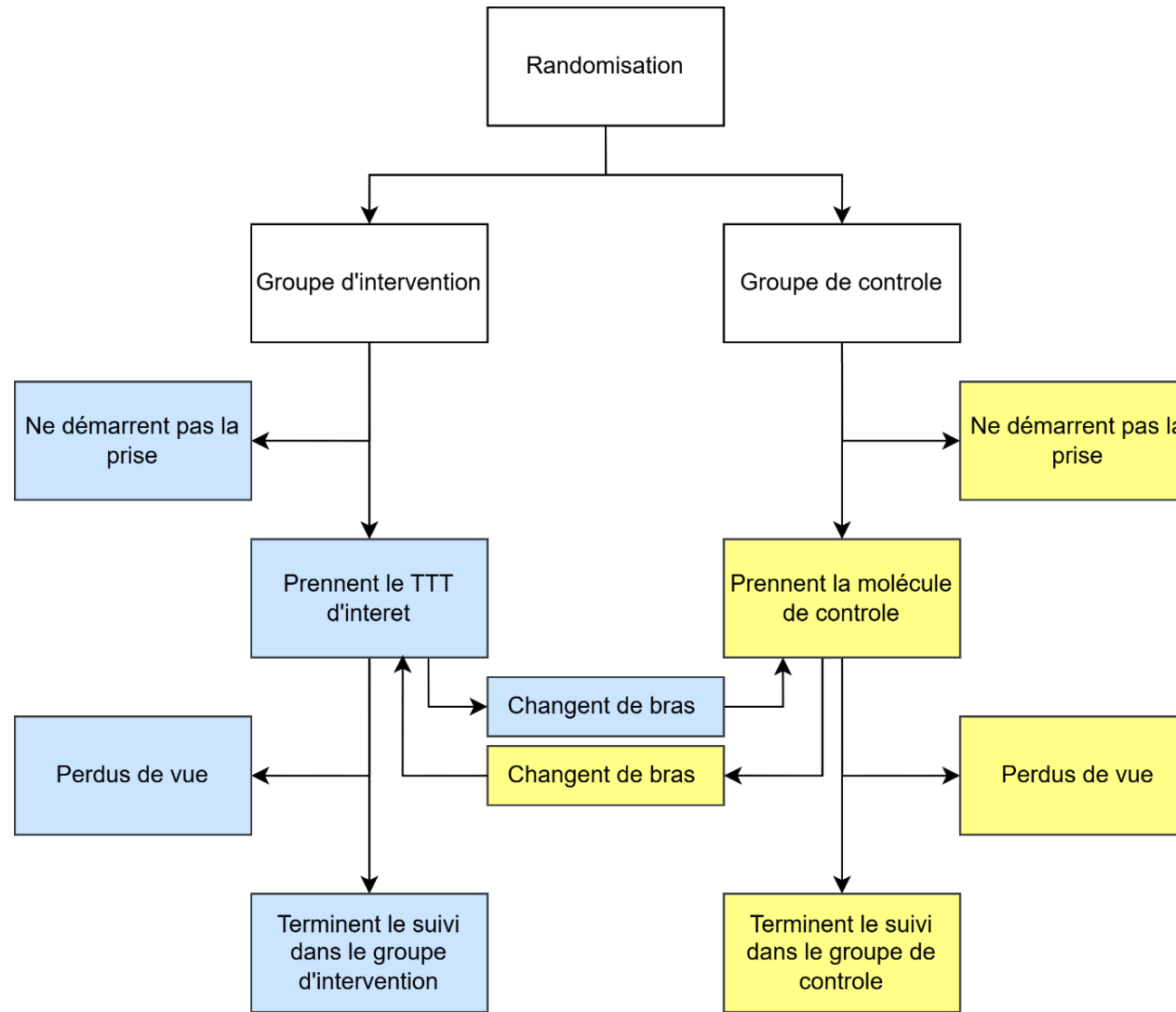
Des patients sont perdus de vue



Imputation des données manquantes

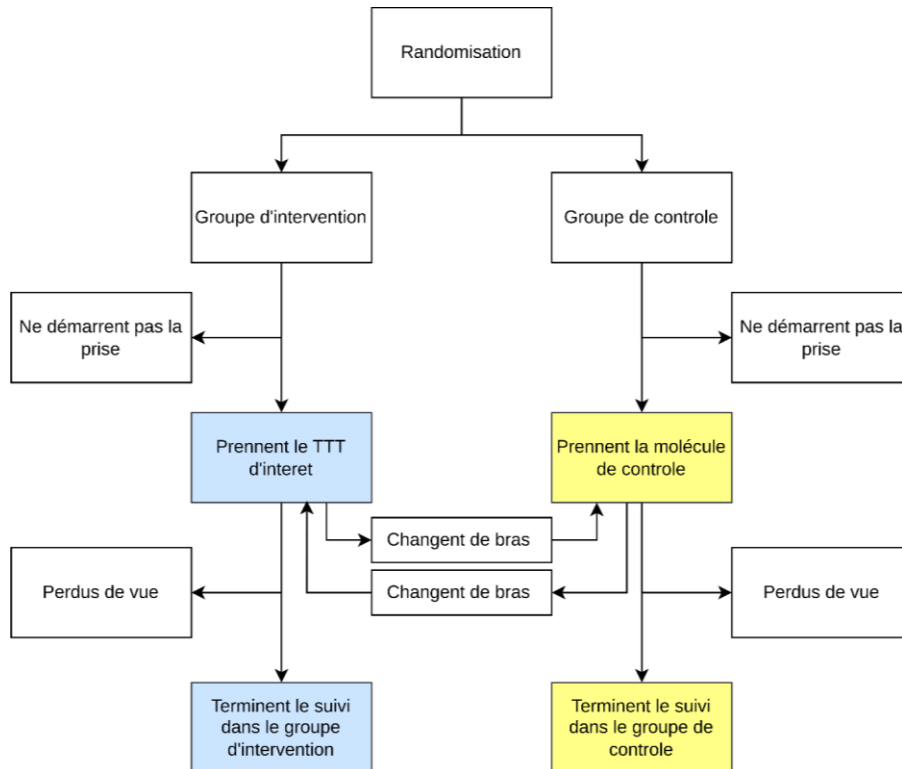
Les (y) absentes doivent être remplacées par des valeurs qui défavorisent l'efficacité du traitement (par exemple, comme si le traitement avait été un échec).

Ensemble d'analyse en intention de traiter (ITT)

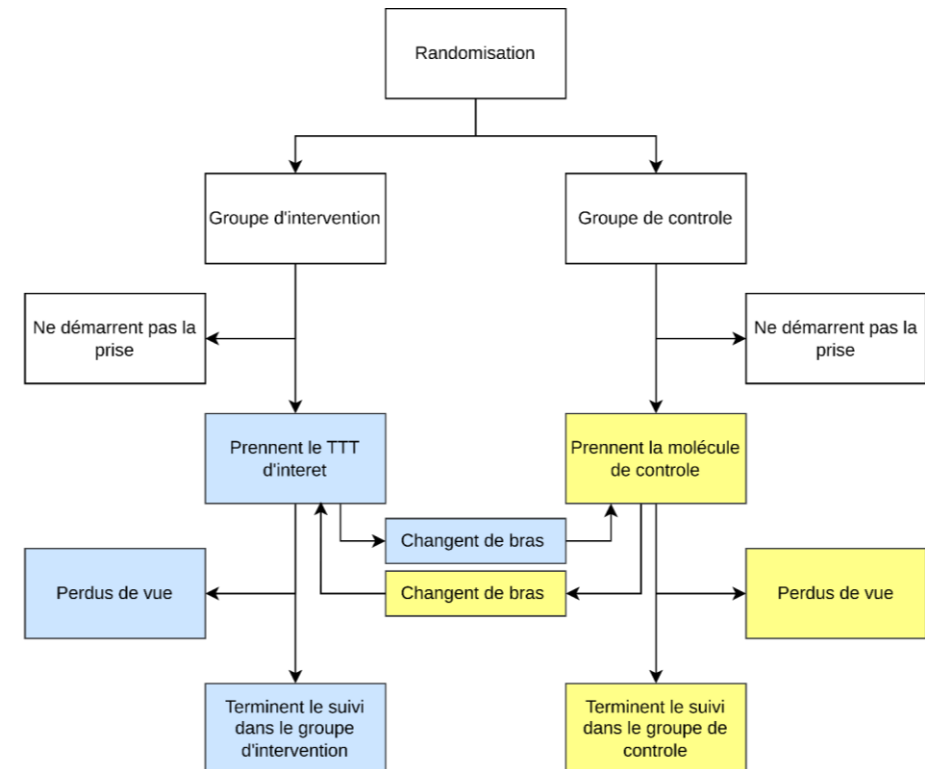


Les autres ensembles d'analyse

Quelle que soit la définition utilisée par les auteurs, **toute analyse qui n'est pas en ITT** doit être considérée comme per protocole par défaut (avec les limites méthodologiques qui suivent).



Un exemple parmi les possibles ensembles per protocole (PP)



Un ensemble modified intention to treat (mITT), dernièrement très à la mode

Impact statistique de l'ensemble d'analyse

- L'analyse **ITT** a tendance à « diluer » la différence entre les deux groupes
 - Dans les essais de **supériorité**, elle est conservatrice : c'est l'idéal !
 - Dans les essais de **non-infériorité**, elle n'est plus conservatrice (car elle rapproche les effets des deux traitements), donc elle n'est plus suffisante à elle seule
- L'analyse **PP** a tendance à « exalter » la différence entre les deux groupes, mais elle est très sujette au biais d'attrition

Attrition et ensembles d'analyse: questions type

- 1. Y-a-t il des données manquantes ? Comment elles ont été gérées ?**
(Risque de biais d'attrition)
- 2. Quel est l'ensemble choisi pour l'analyse principale ?** (Risque de biais d'attrition? Impact statistique ?)
- 3. Si les auteurs n'ont pas utilisé la terminologie standard, définissez l'ensemble d'analyse utilisé.**
- 4. Interprétez la Figure 1.** (flowchart d'inclusion)

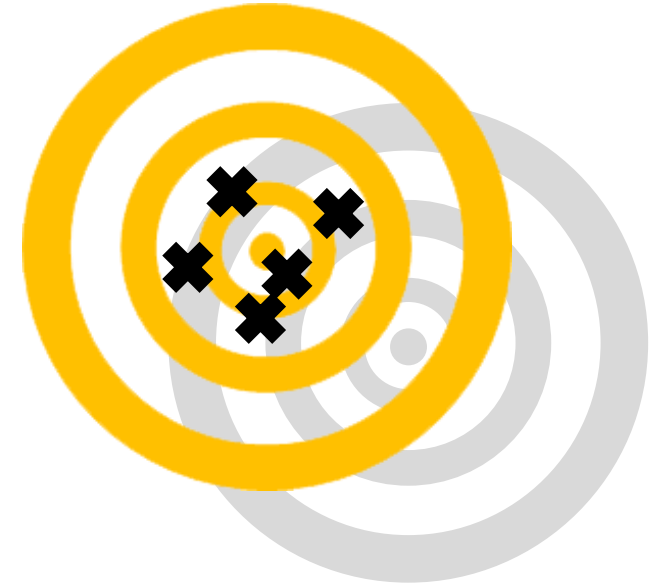
Les biais de classement / mesure

- Biais de **performance** : dans ce qui concerne les soins autres que le traitement étudié, les médecins peuvent prendre en charge différemment les patients des deux bras (par exemple, ils pourraient s'occuper avec plus d'attention des patients du groupe intervention)
- Biais de **suivi** : un suivi plus intense dans un bras peut dénicher plus d'évènements (par exemple, ils peuvent demander plus d'examens biologiques pour les patients du groupe intervention).
- Ces biais sont particulièrement dangereux lorsqu'ils sont **différentiels** entre les groupes !

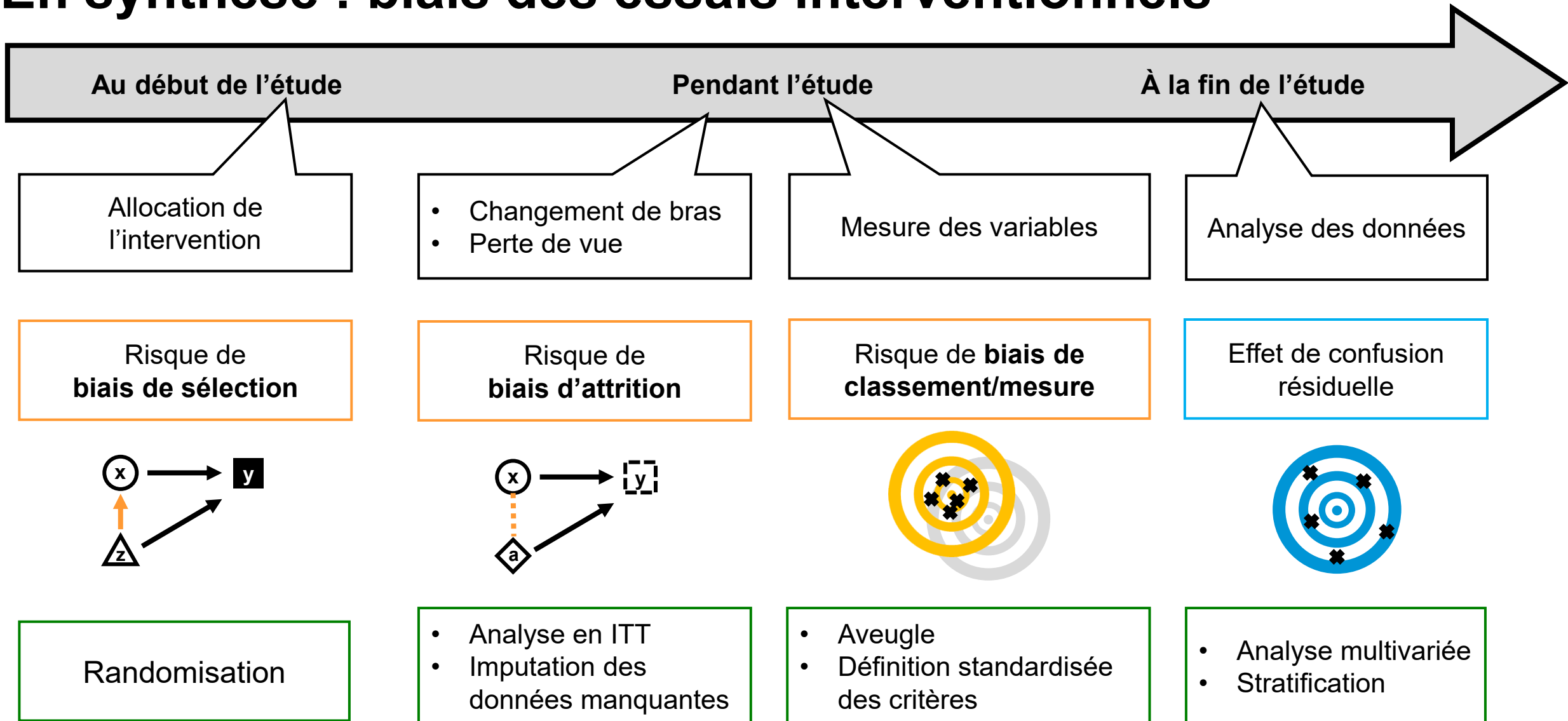
Outils pour prévenir ces biais :

- Réalisation de l'essai en double aveugle
- Définitions standardisées et critères diagnostiques fixes pour adjudiquer les évènements

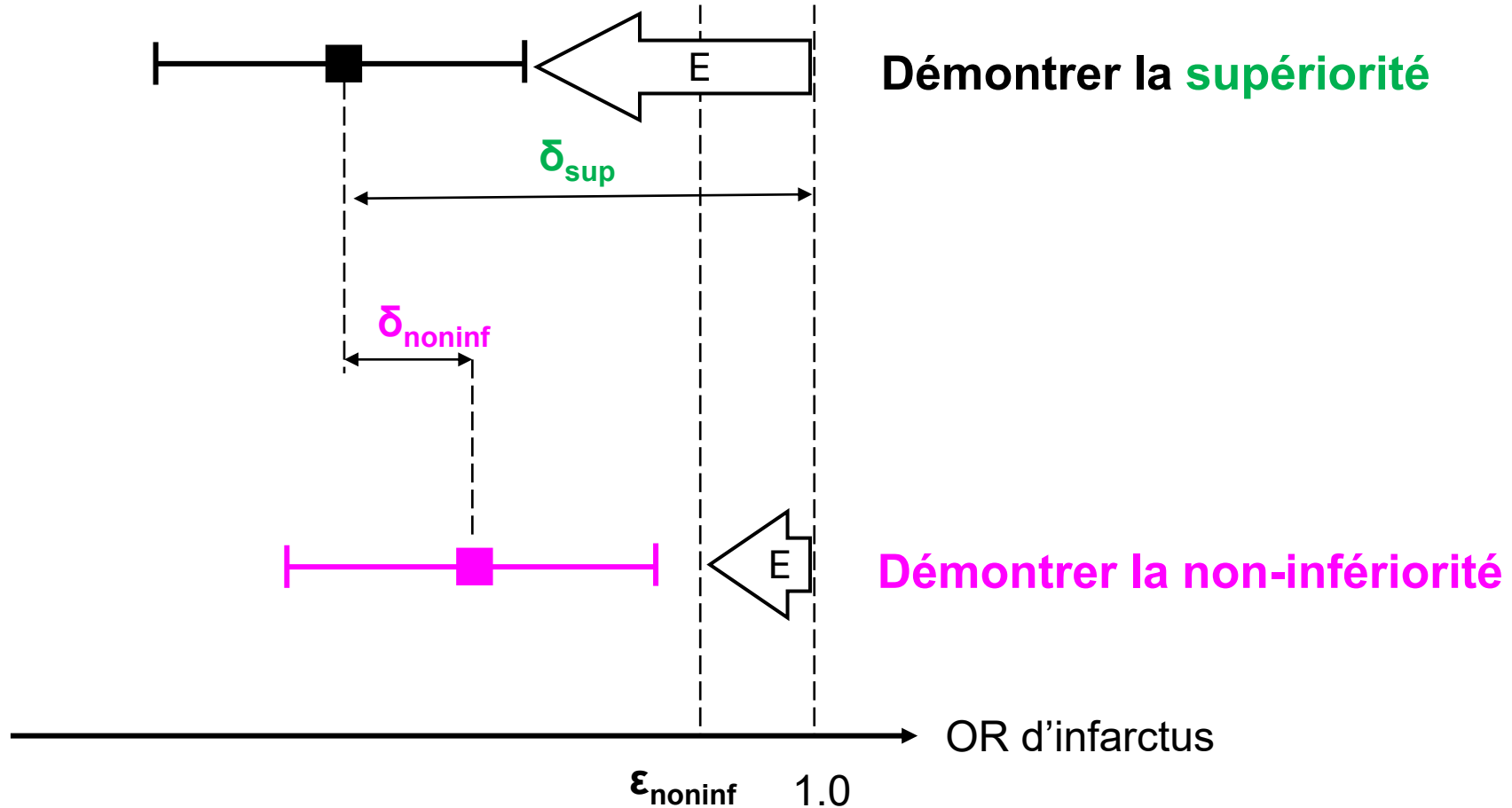
Si l'investigateur ne peut pas être en aveugle pendant la prise en charge, désignation d'un comité d'adjudication des évènements à posteriori, qui peut être en aveugle



En synthèse : biais des essais interventionnels



Objectifs des évaluations d'intervention



Essais de non-infériorité

- Le **calcul du seuil** de non-infériorité sur l'échelle de la mesure d'association (ϵ_{noninf}) doit être détaillé dans les Méthodes
- Par rapport à un design de supériorité, à parité de risque de premier type, la **puissance** d'une étude de non-infériorité est moindre (car $\delta_{\text{noninf}} < \delta_{\text{sup}}$)
- Au contraire que dans les études de supériorité, dans les études de non-infériorité **l'analyse en ITT est trop optimiste** (car elle dilue la différence entre les traitements). Il faut donc prendre en compte aussi les résultats d'un ensemble d'analyse per protocole.

Essais de non-infériorité : quand ?

- Si l'on dispose d'informations **préliminaires** sur l'intérêt d'un deuxième médicament (moins cher, moins d'EI, facilité d'administration...)
- Ces caractéristiques ne doivent pas être évaluées dans l'objectif principal de l'essai de non-infériorité, car le CDJP de celui-ci est **l'efficacité**

Essai de non-infériorité : questions type

- 1. Quels sont les résultats préalables qui justifient un essai de non-infériorité ?**
- 2. Quelles sont les données préliminaires qui étayent le calcul de la borne de non-infériorité ?**
- 3. L'intervention est non-inférieure au comparateur. Alors est-elle supérieure ? (Spoiler: NON, sauf si d'autres analyses spécifiques ont été réalisées)**

Choix du critère de jugement

PERTINENCE

Il faudrait privilégier, si possible, un CDJ

clinique : proche aux questions que le médecin se pose devant ses vrais patients.

- Le patient a moins mal ? Le patient va mieux ?
- Le patient guérit ?
- Le patient décède ?
- Y-a-t il des effets indésirables du traitement ?

OBJECTIVITÉ et REPRODUCIBILITÉ

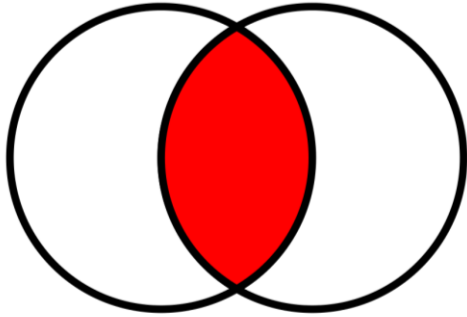
Il faut prévenir dès la phase de rédaction du protocole le risque du biais de subjectivité de l'expérimentateur, en choisissant des CDJ « **durs** »:

- Scores de sévérité validés
- Mortalité toute cause confondue
- Adjudication des événements par un comité (mieux encore si indépendant)

Critères de jugement intermédiaires

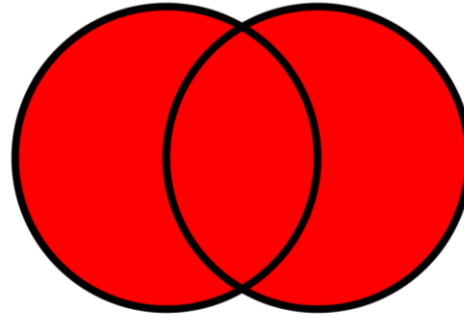
- Parfois, des paramètres biologiques sont utilisés en tant que CDJ : il s'agit de CDJ **intermédiaires**
- Si on veut s'en servir pour tirer des véritables conclusions sur la pathologie de fond (en les considérant en tant que vrais CDJ **de substitution**), on doit pouvoir référencer des études de validation **préalables**
- Une seule et même étude ne peut pas à la fois valider un CDJ intermédiaire en tant que CDJ de substitution ET tirer des conclusions sur la pathologie à partir des mêmes données

Critères de jugement multiples



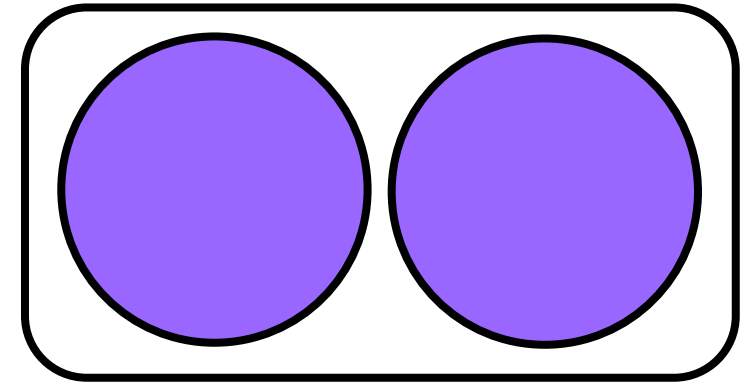
**CDJ co-primaires
(opérateur AND)**

- Plusieurs conditions doivent se vérifier au même temps
- Réduit la puissance de l'étude
- Pas d'inflation du risque alpha
- Exemple : effet sur la douleur **ET** sur la photophobie d'un TTT pour la migraine



**CDJ composite
(opérateur OR)**

- Au moins une condition doit se vérifier
- Augmente la puissance de l'étude
- Pas d'inflation du risque alpha
- Exemple : réduction du risque de AVC, infarctus, **OU** mort, avec un antiplaquettaire



Famille de CDJ

- Au moins une condition doit se vérifier
- N'influencent pas la puissance de l'étude
- Inflation du risque alpha
- Exemple : effet sur la vitesse de guérison, **OU** sur le taux de guérison sans cicatrice, d'un TTT pour les brûlures

Inflation du risque d'erreur de premier type

Le risque de conclure à **tort** qu'il existe un effet du traitement (là où le traitement n'a pas vraiment d'effet) **augmente** quand **plusieurs tests** statistiques essayent de répondre à la même question de fond, et on veut que la réponse positive à un seul d'entre eux **suffise à confirmer l'hypothèse entière**.



*Si **au moins un** des couleurs de bonbons provoque l'acné, alors **les** bonbons provoquent l'acné (hypothèse conjointe)*

Famille de CDJ multiples

Inflation du risque d'erreur de 1^{er} type



*Si **un** des couleurs de bonbons provoque l'acné, alors les bonbons **de tel couleur** provoquent l'acné (hypothèse individuelle)*

Il y a 6 hypothèses, il faudrait réaliser 6 essais cliniques séparés

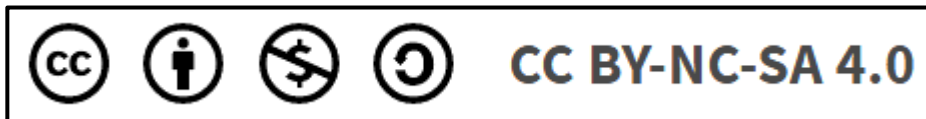
Pas d'inflation du risque d'erreur de 1^{er} type

Critères de jugement : questions type

1. **Le CDJP est pertinent pour répondre à la question de recherche ?**
2. **Le CDJP est suffisamment objectif ?** (Risque de biais de classement/mesure)
3. **Le CDJ intermédiaire a été déjà validé auparavant ?**
4. **Les CDJ multiples sont analysés en tant que composites ou co-primaires ?** (Y-a-t il une inflation du risque d'erreur du premier type?)
5. **S'il y a une inflation du risque d'erreur, est-elle prise en compte par les auteurs avant de tirer des conclusions sur l'hypothèse de recherche ?**



D'AUTRES QUESTIONS ?



LCA des études interventionnelles © 2026 by Giuseppe BALICE is licensed under Attribution-NonCommercial-ShareAlike 4.0 International. To view a copy of this license, visit <http://creativecommons.org/licenses/by-nc-sa/4.0/>