

Quelles données pour quelle recherche en santé

Philippe SAINT PIERRE et Nicolas SAVY
Institut de Mathématiques de Toulouse

Les cafés de l'IMT et fête de la sciences 2021

Le 05 octobre 2021

Contenu de la présentation

Principes généraux de la recherche médicale

Les essais cliniques interventionnels

Les enquêtes observationnelles

La « médecine du 21ème siècle »

- Essai clinique simulé
- Essai clinique émulé
- Jumeau numérique

Principes généraux de la recherche médicale

La recherche a
besoin de temps!

on s'en fout de
la recherche: on
veut un vaccin!



Xavier GANCE -

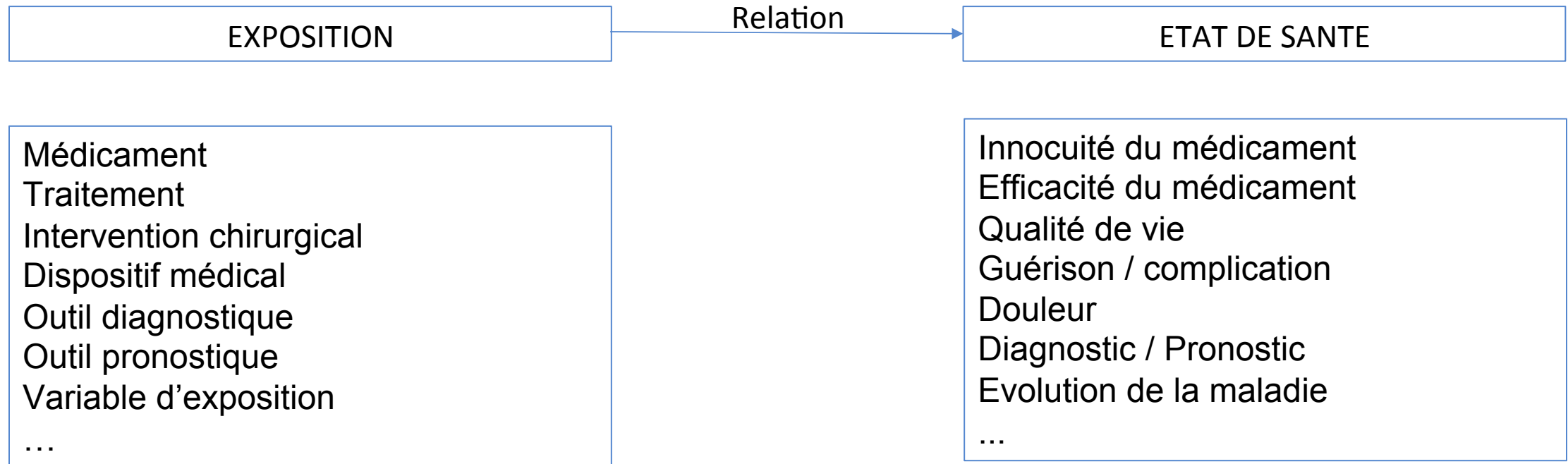
Médecine au XVIII^e siècle



"Les médecins administrent des médicaments dont ils savent très peu, à des malades dont ils savent moins, pour guérir des maladies dont ils ne savent rien"

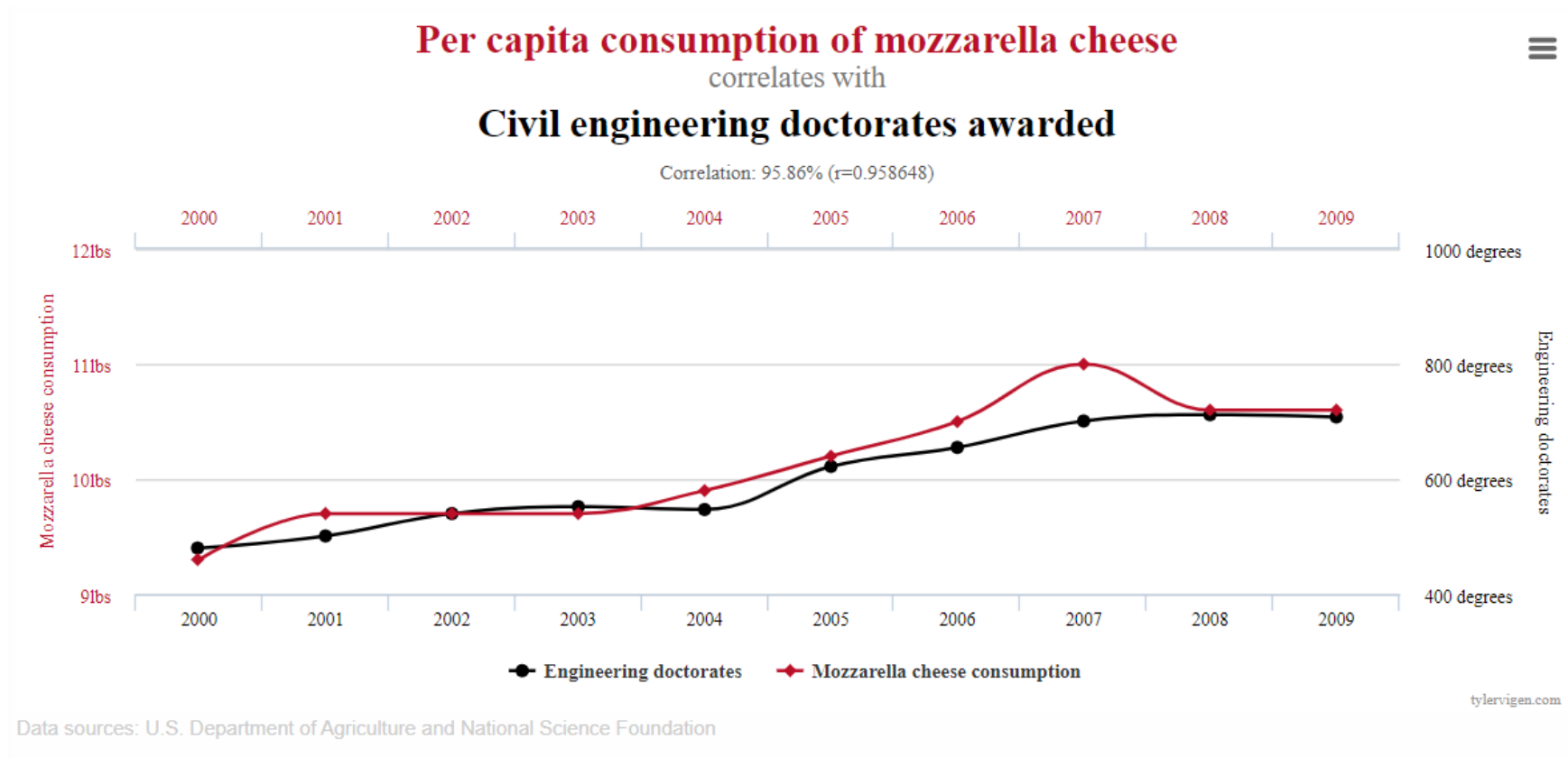
(Voltaire)

Médecine Factuelle (années 1950)



Objectif : Montrer que la relation est **causale**.

- Une approche purement statistique ne suffit pas
- **Association n'est pas forcément causale**



Les **critères de Bradford-Hill**.

Groupe de conditions minimales pour fournir une preuve adéquate d'une relation causale entre deux évènements

Principe de la causalité - Platon

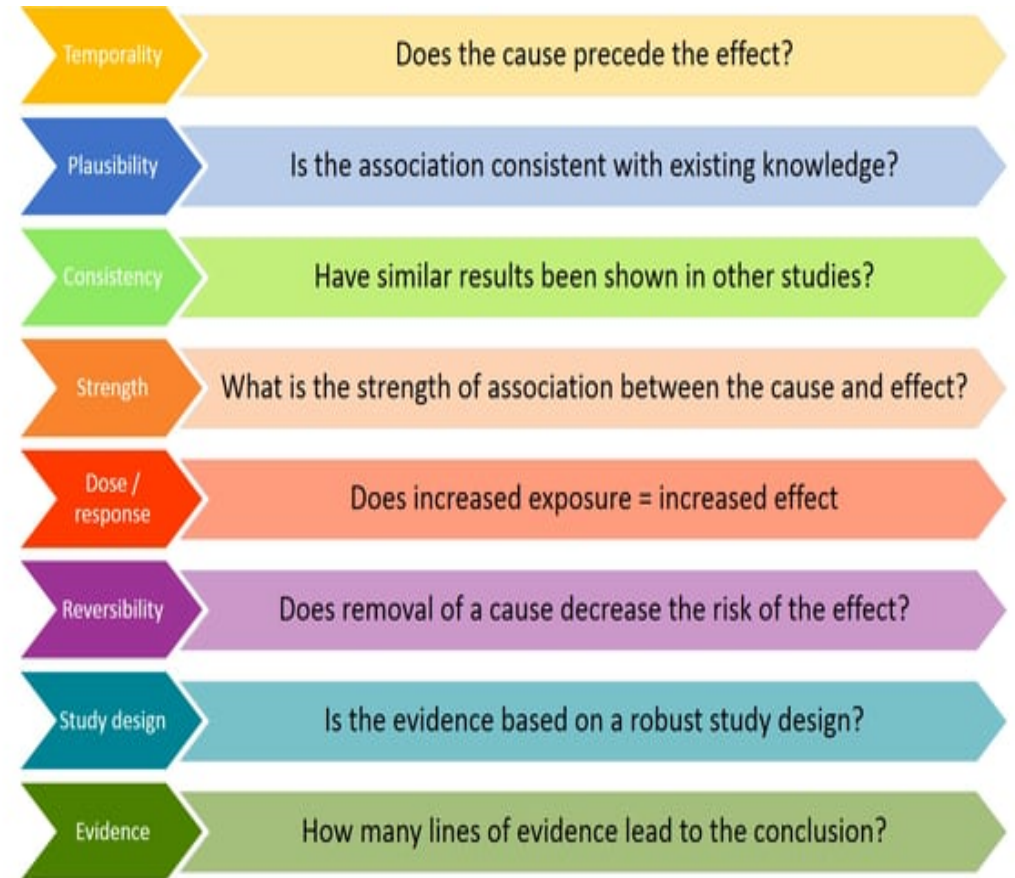
"Sans **l'intervention** d'une cause, rien ne peut être engendré"

Approche **méthodologique** des essais cliniques basé sur un protocole d'étude spécifiant :

- Mode de recueil des données
- Déroulement de la recherche
- Méthode d'analyse des données

Les études interventionnelles naturellement plus performantes

Le « Gold Standard » est **l'essai contrôle randomisé**



Les essais cliniques interventionnels



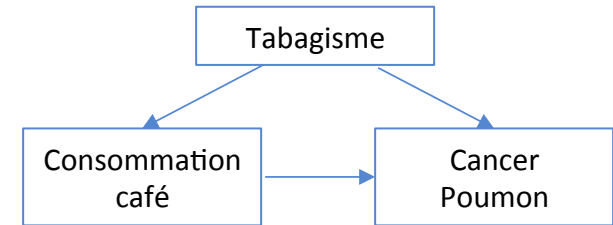
Qualité d'une recherche (médicale)

Tout est question de **biais**...

- Erreur **méthodologique** pouvant conduire à une sur- (ou une sous-) estimation du paramètre étudié
- **Toute étude présente des biais**, il faut avant tout essayer de les contrôler
- Les biais ne suppriment pas la valeur de l'étude, mais **limitent les conclusions**

- **Le biais de sélection**
 - Dû aux procédures de recrutement des sujets de l'étude

- **Le biais de confusion**
 - Dû à la prise en compte (ou non) de facteurs de confusion
 - Facteur lié à l'exposition et à la pathologie modifiant la relation entre ces deux variables



- **Le biais d'information (de mesure)**
 - Dû à l'utilisation d'informations erronées (exposition, diagnostic)
 - Biais de subjectivité
 - Biais de déclaration
 - Biais de mémorisation
 - Biais de classement

- **Le biais d'attrition**
 - Dû aux patients qui sortent de l'étude avant la fin ou à des interruptions de traitement

Pourquoi les essais cliniques contrôlés randomisés Gold Standard ?

Car permet le contrôle « **by design** » de ces biais

- **Le biais de sélection**

- Contrôler par randomisation

- **Biais de confusion**

- Contrôler par randomisation

- **Biais d'information**

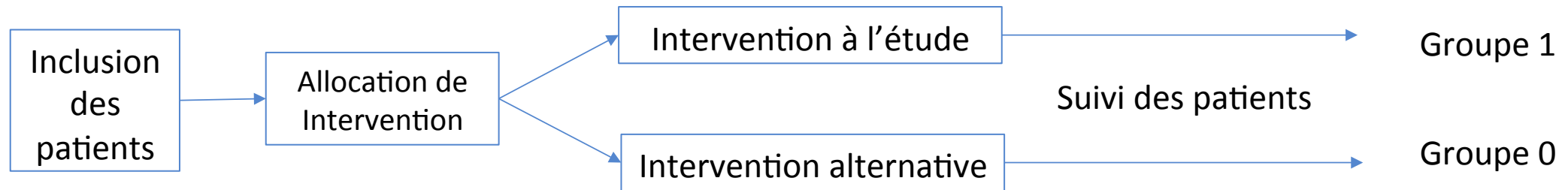
- Contrôler par la mise en aveugle

- **Biais d'attrition**

- Contrôler par « bras contrôle » + hypothèse « Intent to Treat » (ITT)

Pourquoi les essais cliniques contrôlés randomisés Gold Standard ?

Exemple de design : 2 bras parallèles



L'effet moyen de l'intervention à l'étude / intervention alternative est idéalement mesuré par

$$ATE = \mathbb{E}[Y(1) - Y(0)]$$

$Y(1)$ est la réponse à l'intervention pour le bras « intervention à l'étude »

$Y(0)$ est la réponse à l'intervention pour le bras « intervention alternative »

$$\widehat{ATE} = \frac{1}{n_1} \sum_{i=1}^{n_1} Y_i(1) - \frac{1}{n_0} \sum_{i=1}^{n_0} Y_i(0)$$

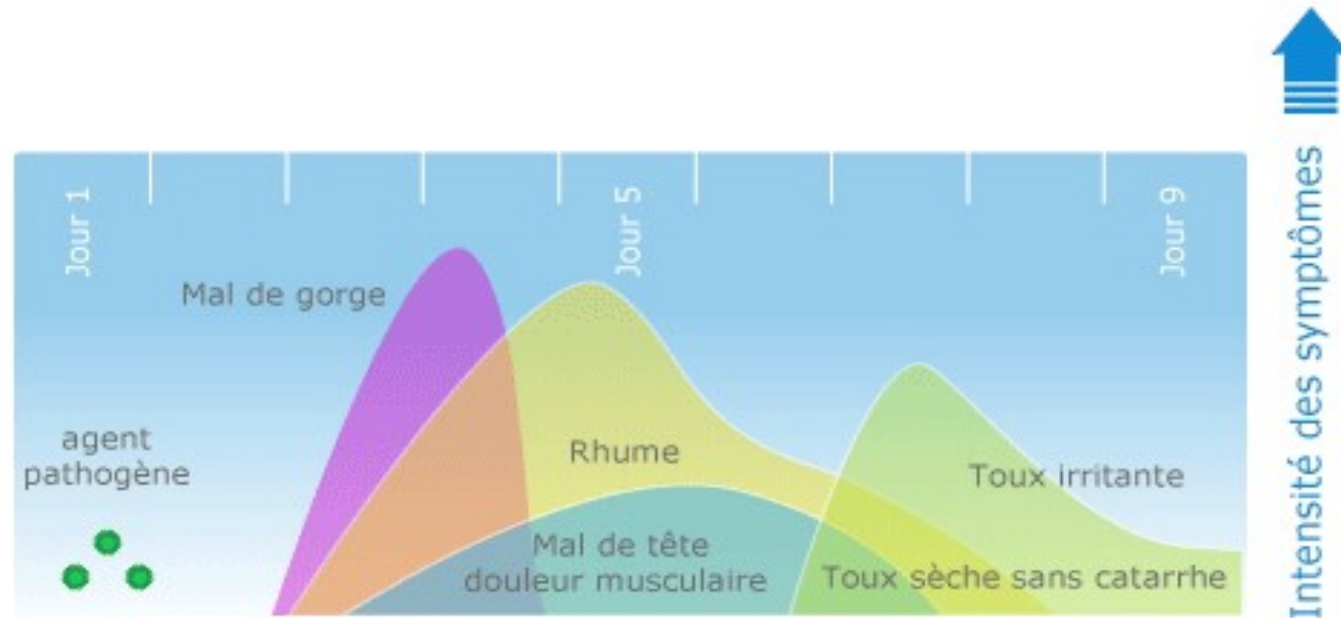
Comparaison des groupes 0 et 1

Intervention alternative : le groupe contrôle

Objectif : Mettre en évidence la différence entre

- ce qui est dû à l'intervention à l'étude et
- ce qui est dû à l'histoire naturelle de la maladie

SCHÉMA D'ÉVOLUTION DES SYMPTÔMES
DE L'INFECTION VIRALE



Pourquoi un groupe contrôle ?



L'art de la médecine consiste à distraire le malade
pendant que la nature le guérit.

(Voltaire)

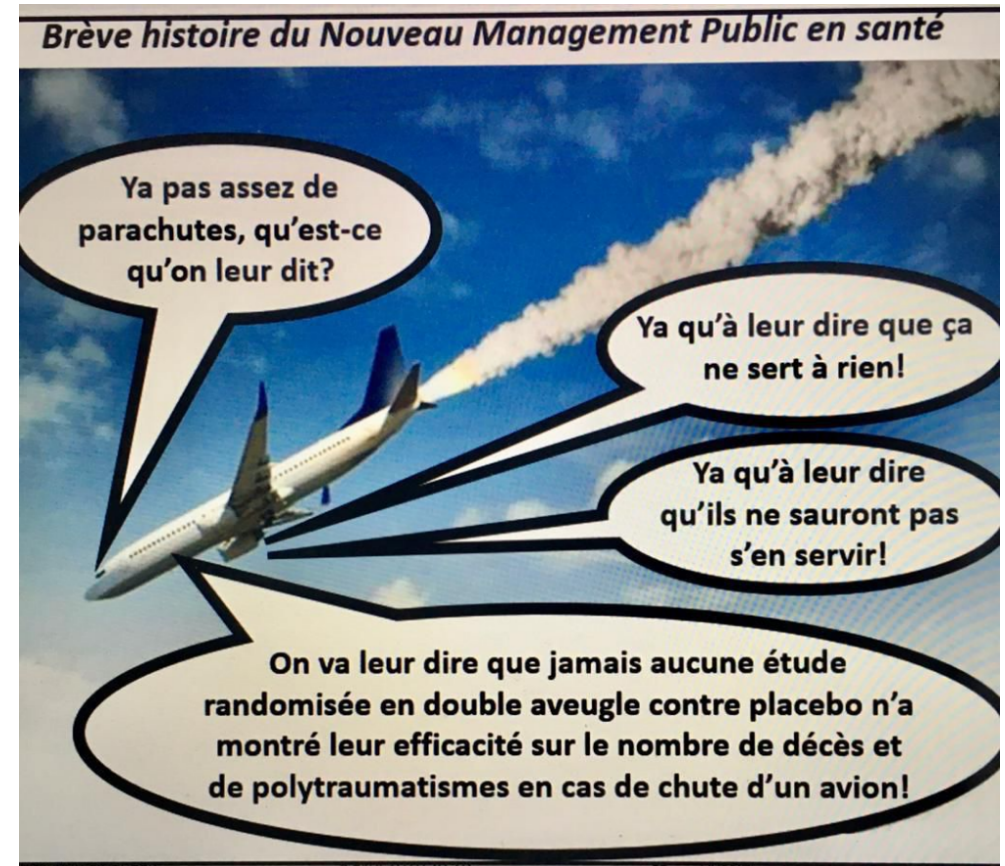
La Randomisation

Allocation de l'intervention dans chacun des groupes est aléatoire.

Pour équilibrer l'influence de facteur sur les bras d'intervention

- Maîtrise du biais de sélection
- Maîtrise du biais de confusion

Il n'est pas toujours possible ni éthique de randomiser !



L'aveugle

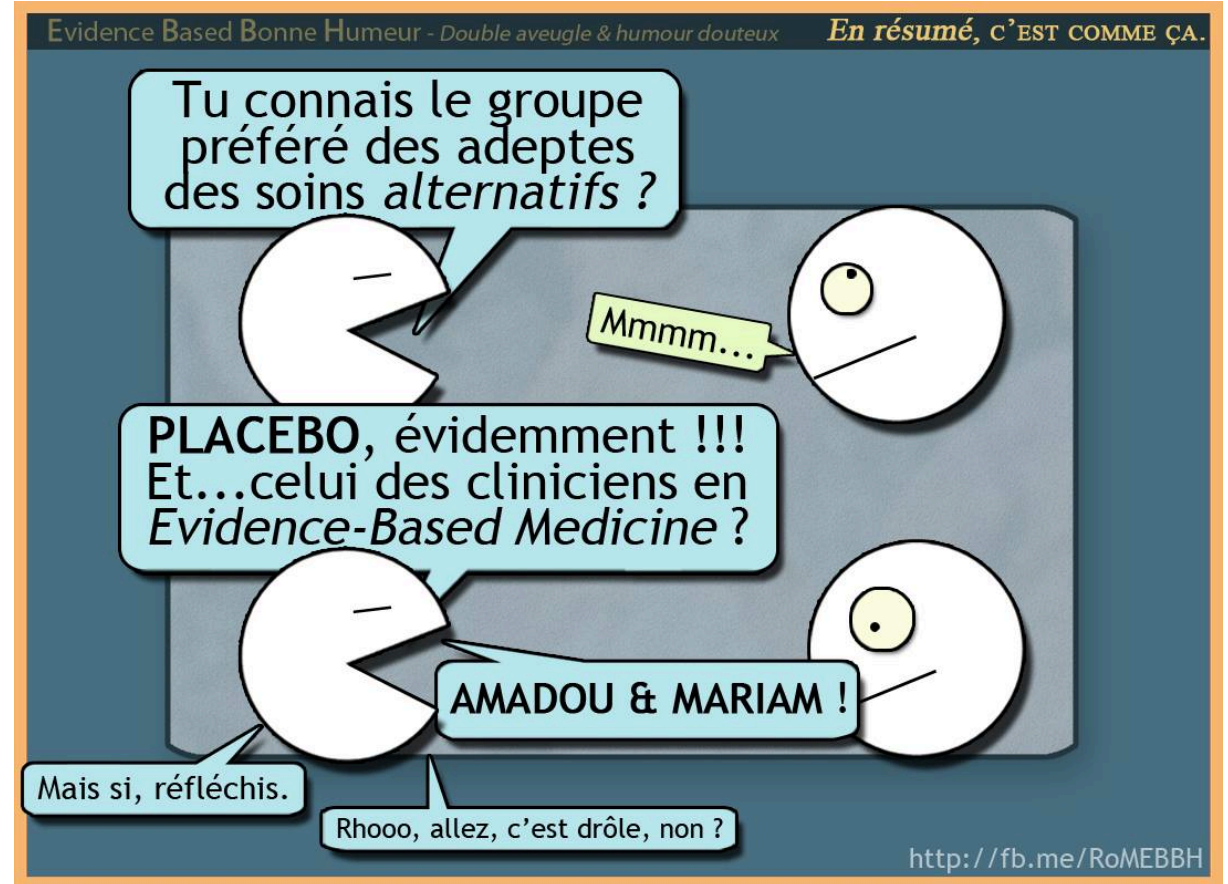
Simple aveugle :

Le patient ne sait pas le bras d'intervention qui lui est attribué

Double aveugle :

Ni le patient, ni le soignant ne sait pas le bras d'intervention qui lui est attribué

Meilleur contrôle du biais d'information

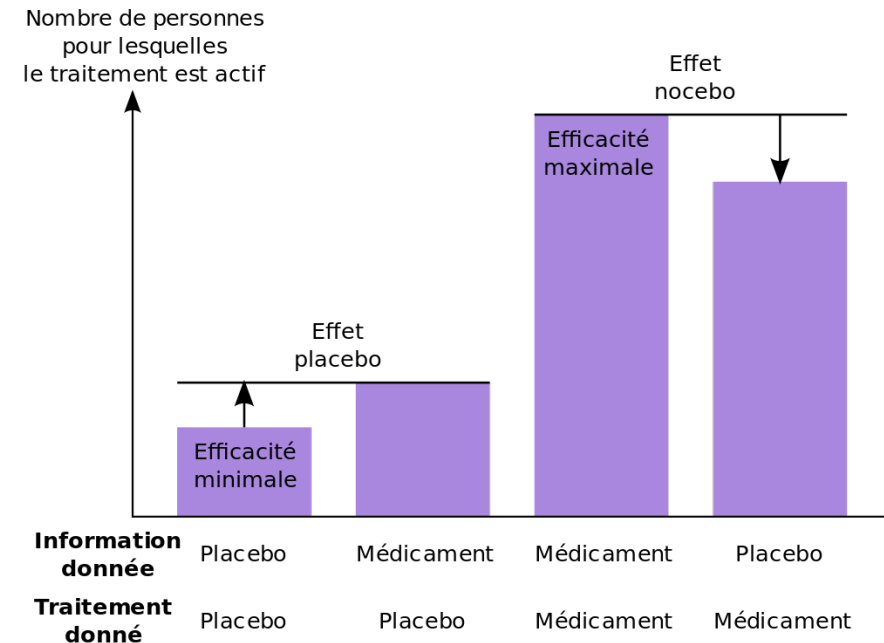


Pourquoi un groupe placebo contrôlé ?

Intervention alternative est une intervention « neutre ».

Médicament : même traitement sans substance active.

Mesure de l'effet propre de l'intervention et non un effet différentiel.



Limites du RCT

Le RCT est le plus haut « niveau de preuve » mais a aussi ses limites !

Sources de variation en recherche médicale

- Variation liée à la maladie
 - Fixer (contrôler) la pathologie
- Variation liée à l'exposition
 - Permet de mettre en évidence l'effet de l'intervention
- Variation liée au patient
 - Inévitable dans un essai sur l'homme

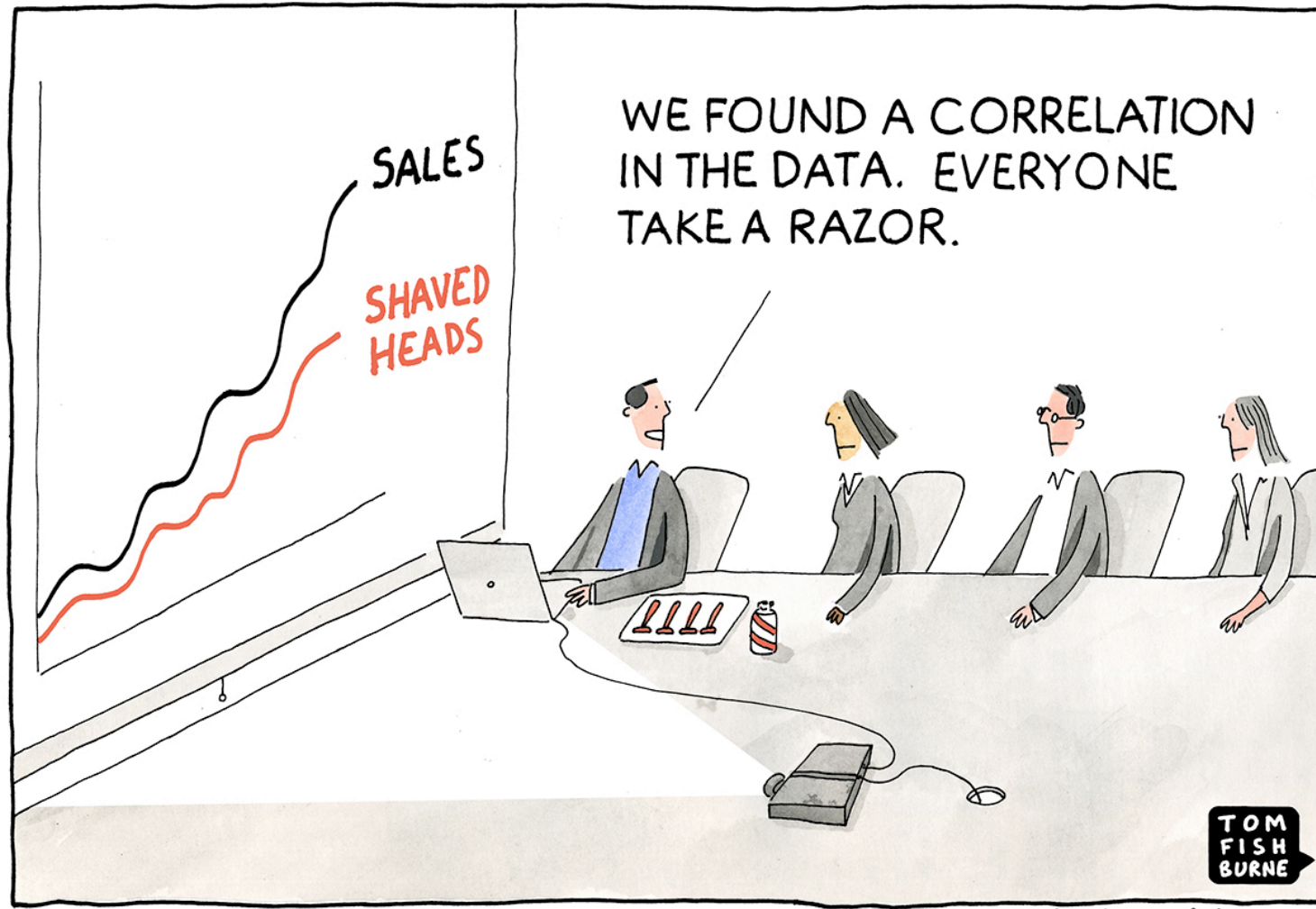
Principe central de l'EVB

Plus la population est **homogène** (critères inclusion / exclusion) plus il est facile de mettre en évidence un effet de l'intervention

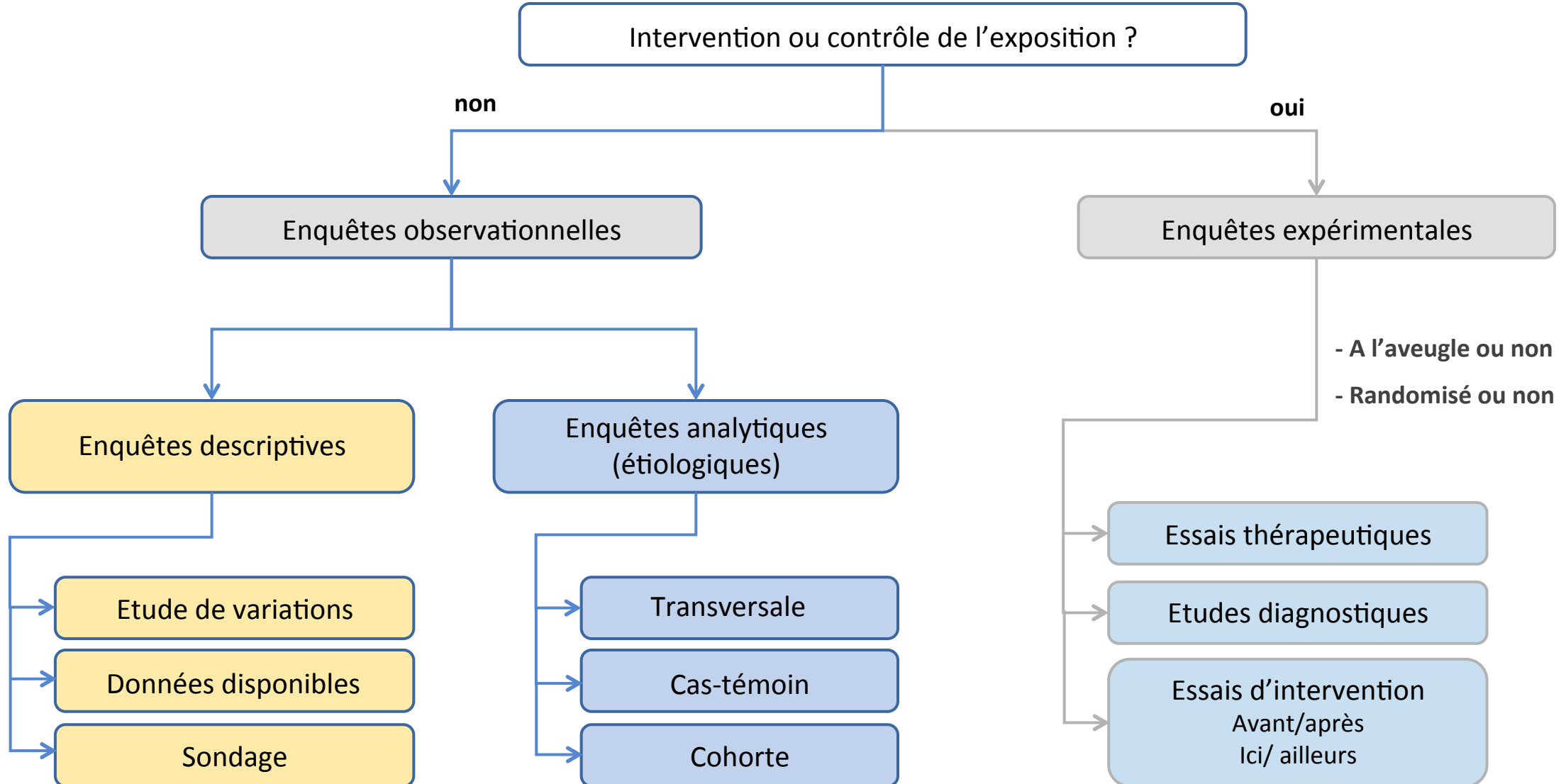
Problème : Validation externe



Les enquêtes observationnelles



Différents types d'enquêtes



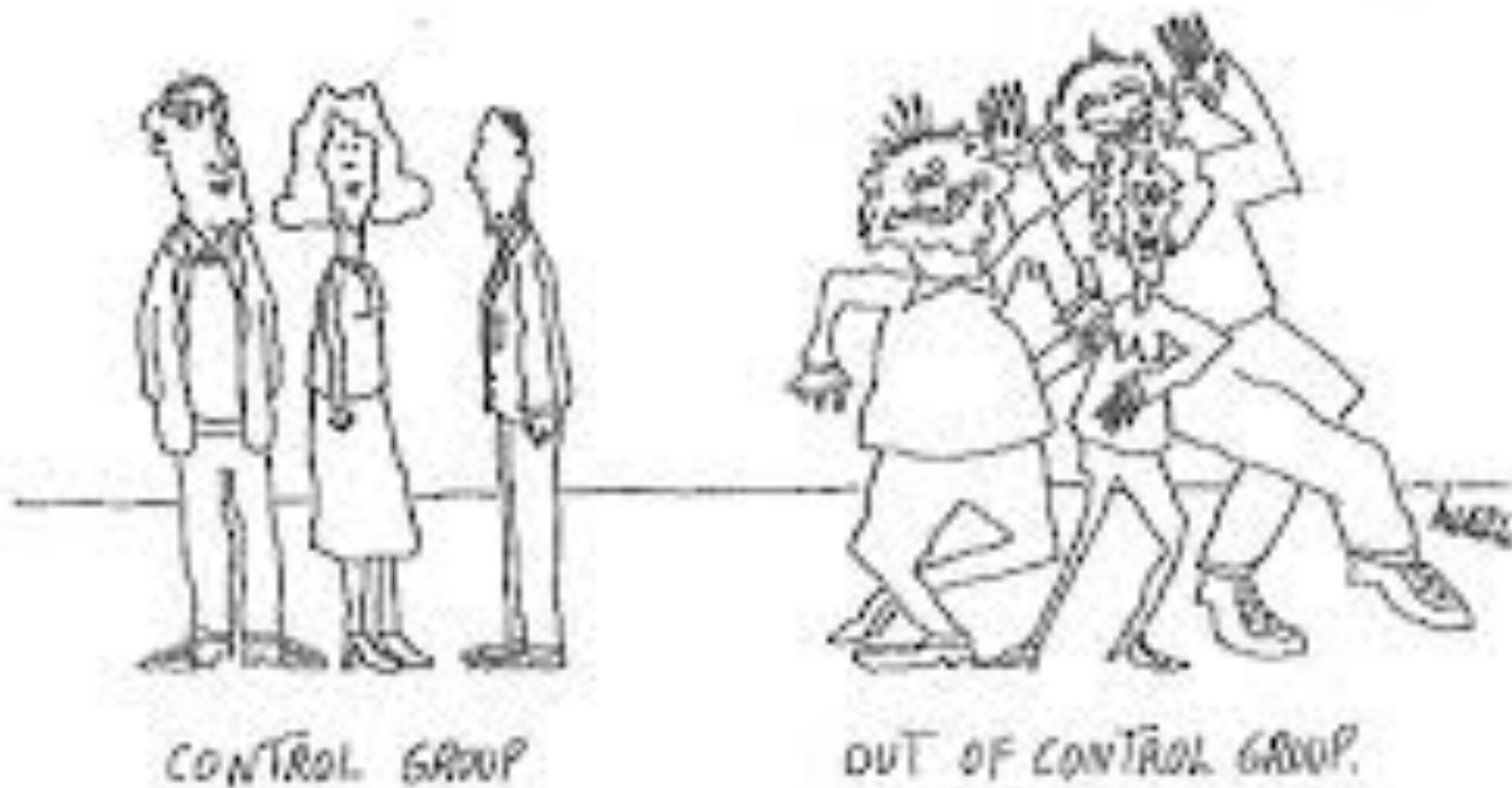
Enquêtes descriptives

- **Statistiques disponibles**
 - **Statistiques vitales** : naissance et causes médicales de mortalité (CepiDc)
 - **Statistiques de morbidité** : registre des cancers, des maladies cardio-vasculaires, ...
 - **Statistiques d'exposition à des facteurs de risque** : expositions professionnelles, ...
 - **Statistiques hospitalières**, programme de médicalisation des systèmes d'information (PMSI)
 - **Déclarations obligatoires** : maladies transmissibles, sida, certificat de santé des enfants, ...
 - **Réseaux sentinelles** de médecins ou de biologistes : grippe, rougeole, coqueluche ...
 - **Enquêtes spécifiques** (de grande envergure pour mieux comprendre un problème de santé : l'importance, les déterminants, ...)
- Études **des variations temporelles** ou **géographiques** des paramètres de santé
- **Enquêtes peu coûteuses et rapides** permettant d'évaluer une hypothèse ou d'alimenter un faisceau d'arguments

Enquêtes analytiques

- Enquête sur un **échantillon de sujets** veillant à recueillir des **informations individuelles** sur la maladie et les expositions
- Etudier la **relation entre la maladie et les expositions** et fournir des arguments en faveur d'une **interprétation causale** (force de l'association, relation dose-effet, chronologie ...)
- Souvent les **seules enquêtes possibles**
- **Enquête transversale**
 - Etude d'un **échantillon représentatif** de la population
 - **Limites** : exposition ou maladie rare, recueil exposition rétrospectif
 - Niveau de preuve le plus faible

Les enquêtes cas-témoins



Enquêtes cas-témoins

- Sélection d'un échantillon **de malade (cas)** et de **non malade (témoins)**
- **Objectif** : comparer la répartition des facteurs de risques entre les cas et les témoins
- Adapté aux **maladies rares**
- **Limites**
 - Recueil de l'information sur l'exposition est toujours rétrospectif (biais de classement)
 - Difficulté à choisir les témoins (biais de sélection)
 - Témoins représentatifs de la population dont sont issus les cas
 - Pas d'estimation de la prévalence (probabilité d'être malade)
 - Pas de suivi dans le temps

Les enquêtes de cohorte

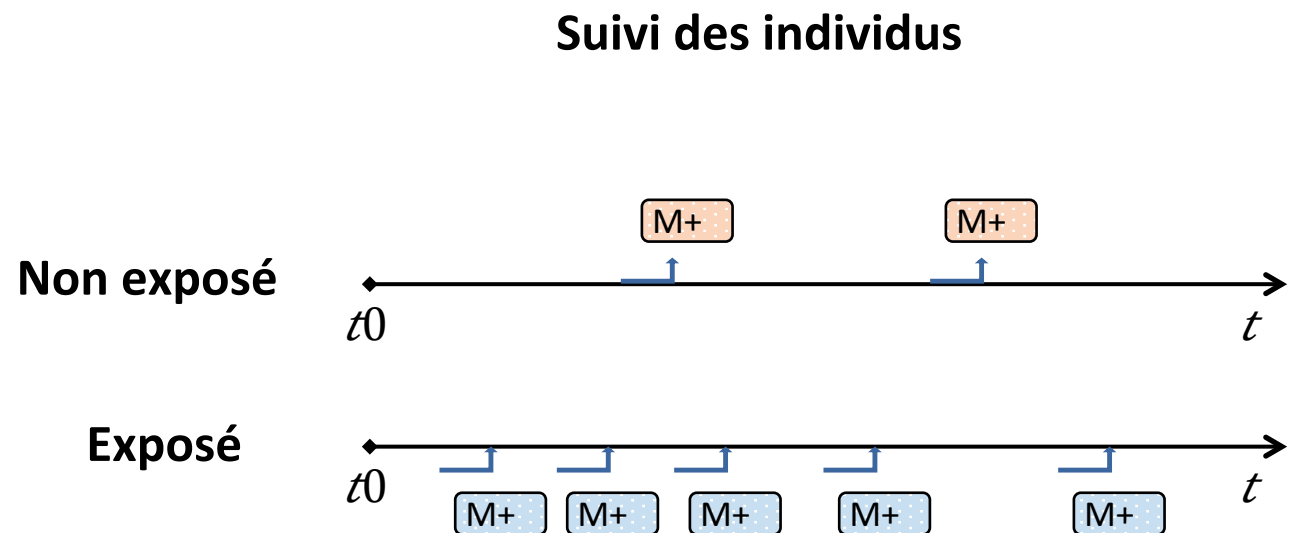


Les légions **romaines** se divisent en dix **cohortes** de six cents légionnaires chacune

Enquêtes de cohorte

- **Suivi dans le temps** (étude longitudinale), selon un protocole pré-établi, d'un groupe de sujets qui répondent à une définition donnée
- Sélection d'un échantillon de **non malades** : **exposés et de non-exposés**

- **Objectif** : comparer l'apparition de la maladie chez les exposés et non exposés



Mise en place d'une enquête de cohorte

- Type de cohorte selon la **chronologie du recueil de données**
 - ↳ Cohorte **prospective**, cohorte **rétrospective**
- Type de cohorte selon de **recrutement des sujets**
 - ↳ Cohorte **fixe**, cohorte **dynamique**, cohorte **mixte**
- Choix de la population source, critère d'inclusions, évaluation des biais de sélection
- Recueil de l'information (Mesures de l'exposition, de la maladie, des facteurs de confusions, ...)
- Modalités de suivi

Enquêtes de cohorte

- **Avantages**

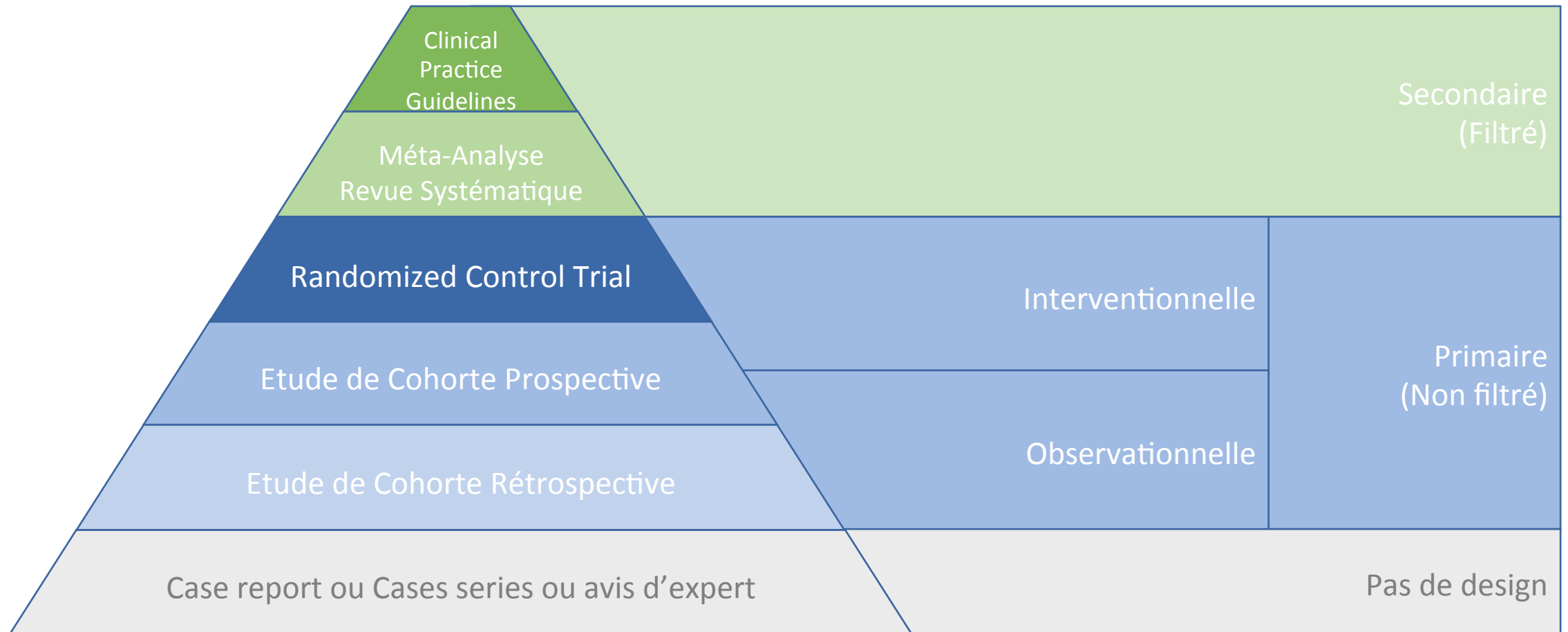
- Permettent d'établir la **séquence chronologique** exposition-maladie
- Recueil prospectif de l'exposition et de la maladie
- **Estimation de l'incidence** de la maladie dans les groupes exposé et non-exposé
- Permettent d'examiner les **conséquences multiples d'une même exposition**
- **Niveau de preuve supérieur** à celui des enquêtes cas-témoins
- **Indispensable pour les études post-AMM**

- **Limite**

- Patients dont le suivi est interrompu : Perdus de vue (biais d'attrition)

« Pyramide » du niveau de preuve

- Le contrôle des biais définit le **niveau de preuve**
- **Discussion** des biais dans la publication



La médecine du XXIème siècle



Assez d'essais !!!

On croule sous les données médicales
Données médicales sous-utilisées

Dans un contexte de « **médecine 6P** »
Dynamique forte vers le **data-reuse**

Essais Simulés

Essais Emulés

Jumeaux numériques

La Médecine du 21^{ème} siècle... une histoire de P...

Contexte :

Avènement de la génomique, de l'épigénétique, du "Big Data« ,... bref beaucoup de données...

Fin du "règne" du tout Méthodologique - Médecine basée sur la fouille de données... « Google Medicine

Personnalisée : Tient compte du profil clinique, génétique, génomique et environnemental **unique** de chaque personne,

Préventive : Prend en considération les problèmes de santé en se concentrant sur le mieux-être **individuel** et non la maladie,

Prédictive : Consiste à utiliser des marqueurs **individuels**, le plus souvent biologiques, pour prévenir, dépister ou traiter les maladies.

Participative : Permet **à chaque patient** de participer à ses propres soins, avec le soutien coordonné de son équipe de soins de santé.

La Médecine du 21^{ème} siècle... une histoire de P...

Changement complet de **paradigme** :

- Plus la population est hétérogène plus ces approches sont pertinentes
- Les données à recueillir sont souvent volumineuses.. voir pire...
- Les méthodes de traitements doivent être adaptées à ce type de données

Il manque donc un P... celui de **Preuve...**

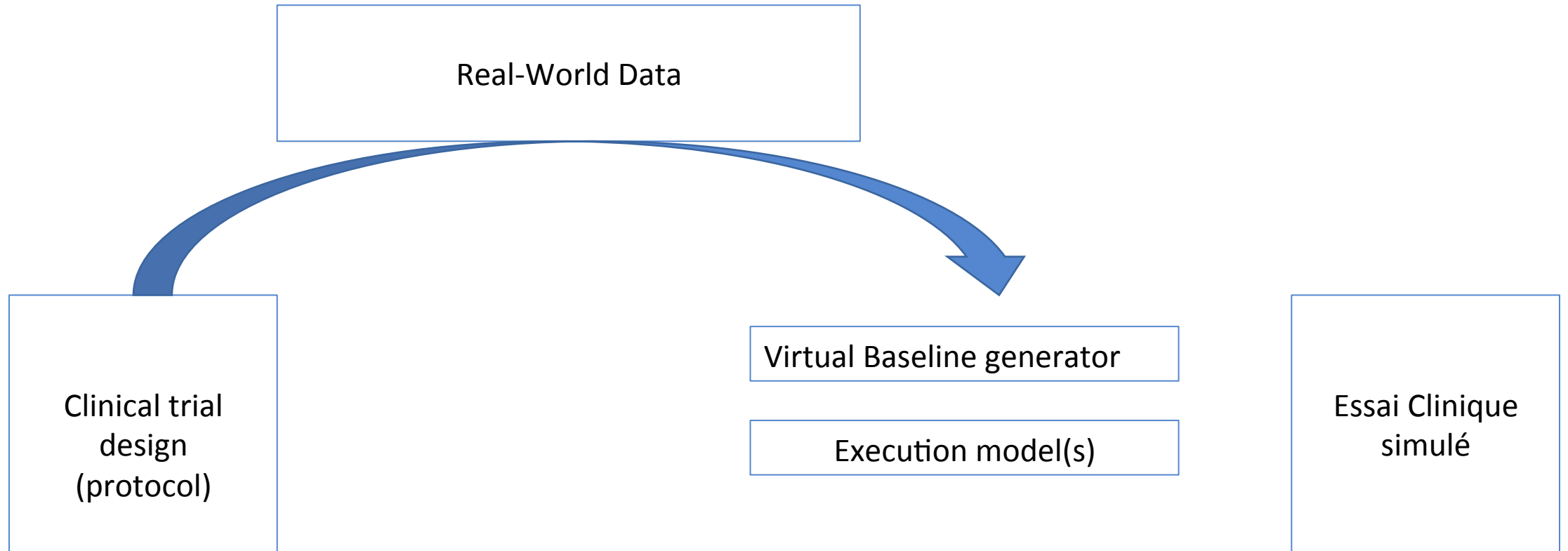
Précision : Fait référence à l'adaptation du traitement aux caractéristiques individuelles de chaque patient.

Ne veut pas dire l'identification d'une intervention **unique pour chaque patient** mais **veut dire** la possibilité de **classer les individus** en **sous-populations** qui diffèrent par leur réponse à une intervention spécifique.

Pour une **sous-population donnée** il est possible de mettre en place une approche "**data driven**" (**Evidence-Based**) d'analyse.

Parcours de soin : Fait référence à la conception de la santé comme un enchaînement d'actes médicaux, on parle de parcours de soin.

La Médecine du 21^{ème} siècle... une médecine virtuelle



Nicolas Savy¹, Philippe Saint-Pierre¹, Stéphanie Savy², Sylvia Julien³ et Emmanuel Pham³

1: Institut de Mathématiques de Toulouse – 2: Estrials - 3 : IPSEN Innovation SAS

“In Silico Clinical Trials”: a way to improve drug development?

Proceedings of JSM 2019 – Biopharmaceutical Section, Denver, 2019.

Virtual baseline Generator

- Modèle de génération de données patients à l'inclusion - on parle de **patients virtuels**
- Simulations des patients virtuels
 - Ces patients suivent une certaine **répartition multivariée** (distribution)
 - **Estimation de la distribution multivariée** à partir de données disponibles et des avis d'experts
 - **Simulation de patients virtuels** qui suivent la distribution estimée
- Patients virtuels sont **similaires** mais ne sont pas des copies identiques
- Permet d'explorer des zones pour lesquelles on a peu de données

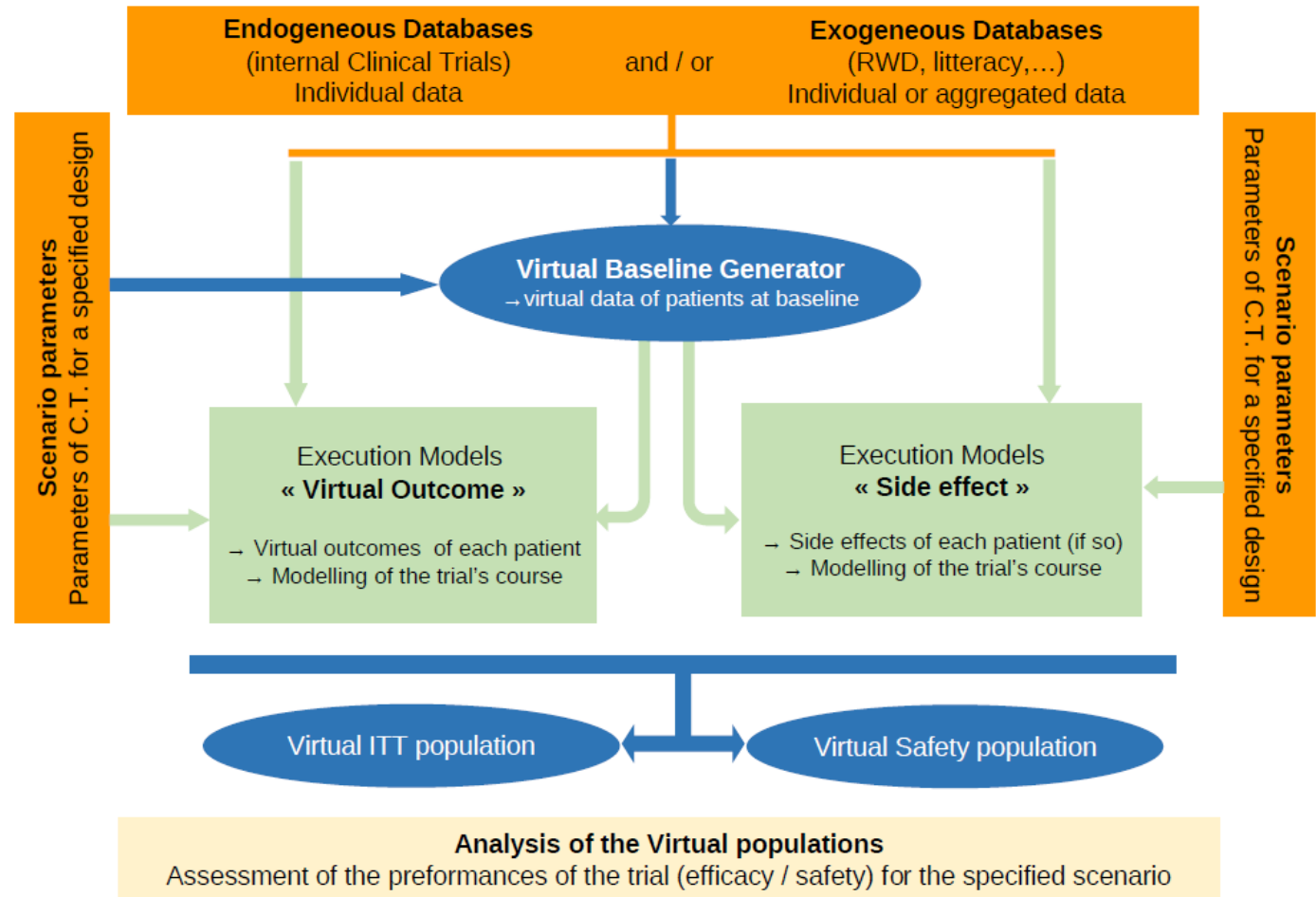
Virtual baseline Generator

- Modèle de génération de données patients à l'inclusion - on parle de **patients virtuels**
- Simulations des patients virtuels
 - Ces patients suivent une certaine **répartition multivariée** (distribution)
 - **Estimation de la distribution multivariée** à partir de données disponibles et des avis d'experts
 - **Simulation de patients virtuels** qui suivent la distribution estimée
- Patients virtuels sont **similaires** mais ne sont pas des copies identiques
- Permet d'explorer des zones pour lesquelles on a peu de données
- Patients virtuels construits sur des modèles purement mécanistiques (simulation des différents organes)



Modèles d'exécutions : évolution temporelle des patients virtuels

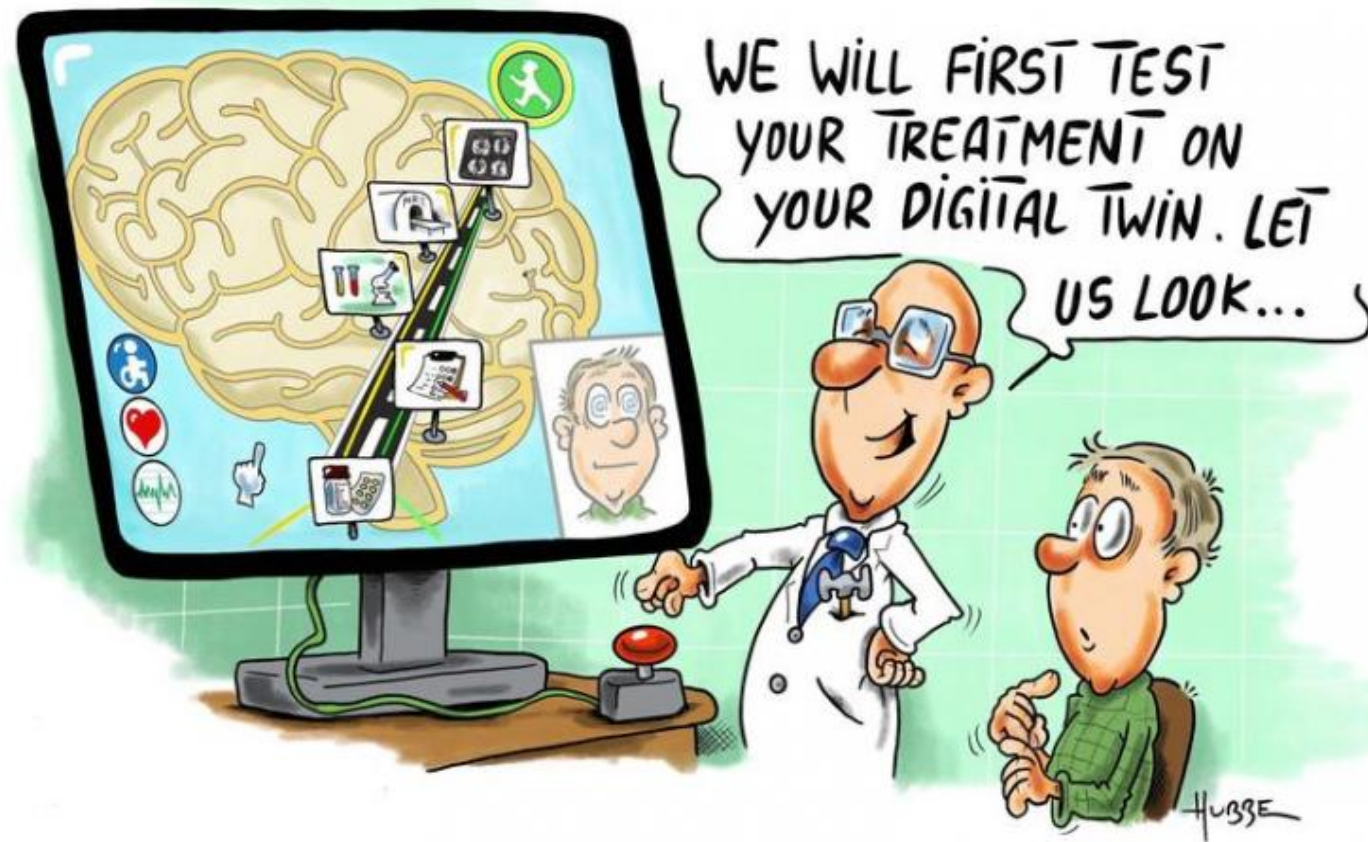
- Evolution de la maladie (outcome)
- Effet du médicament
- Sortie d'étude
- Compliance
- Effets secondaires
- Qualité de vie
- ...



Pour quoi faire ?

- **Essais cliniques simulés (in silico clinical trials)**
 - Encouragé par la FDA
 - Exploitation des données de vie réelles (RWD) disponibles (US, Angleterre)
 - Développement des méthodes de machine learning et d'intelligence artificielle
- **Avantages**
 - Le même patient virtuel peut tester différents médicaments (protocoles ou scénarios)
 - ⇒ On peut mesurer l'Average Treatment Effect (ATE) pour chaque patient (virtuel)
 - Considération éthique : identifier les patients qui ne répondent pas au traitement
 - Considération économique : gain de temps, moins de patients
- Outil pour optimiser ou **pré-calibrer le design** des essais cliniques randomisés

Jumeau numérique



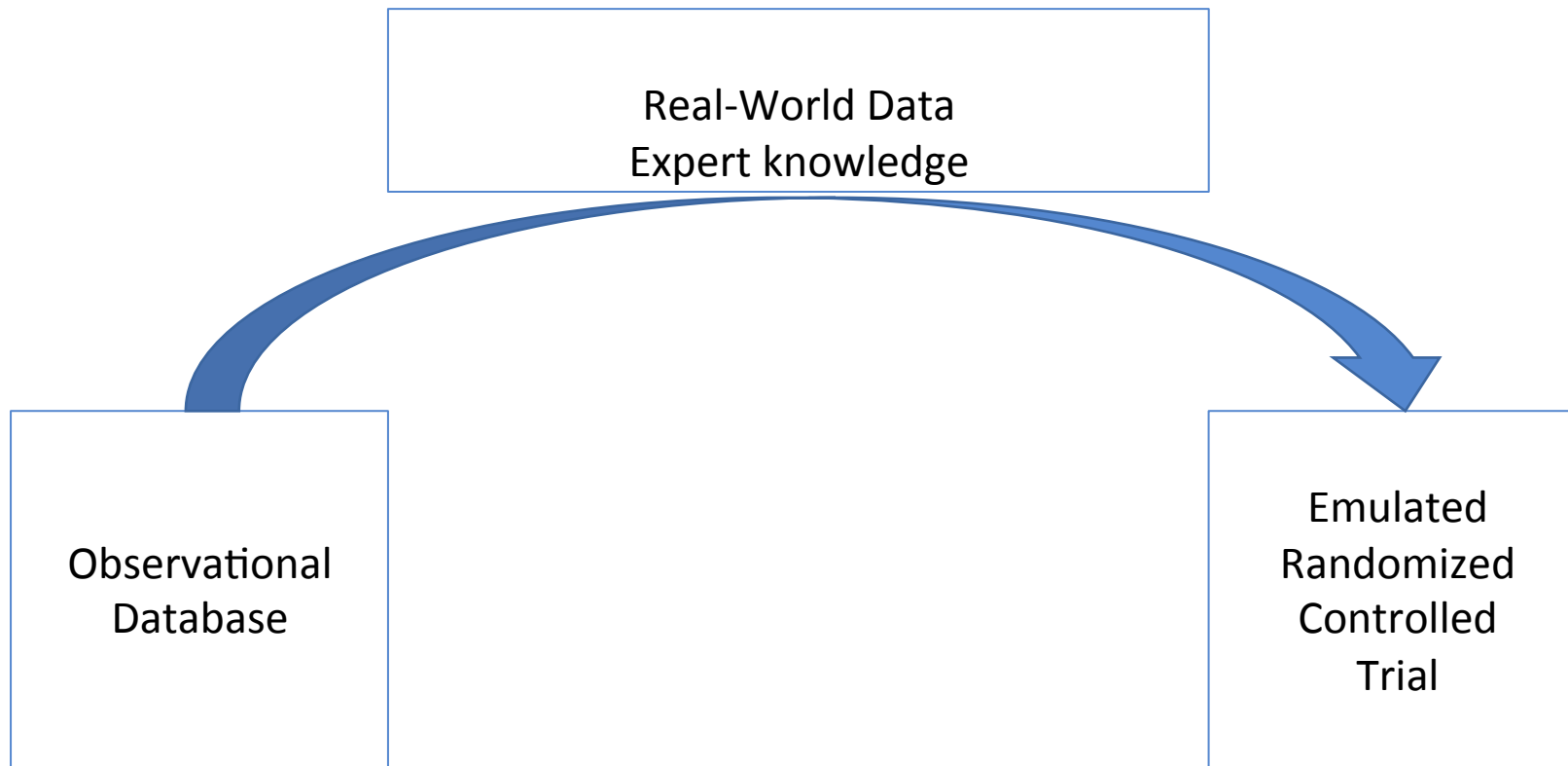
Jumeau numérique

- Créer une version digitale d'une chose réelle
- Permet de tester différents scénarios sur la version digitale
- Nombreuses applications industrielles
- Essai clinique simulé où seule l'outcome est simulée



Essais cliniques émulsés





Essai clinique simulé avec un virtual baseline Generator particulier qui consiste en un « redressement » d'une base de données observationnelle.



Am J Epidemiol. 2016 Apr 15; 183(8): 758–764.

Published online 2016 Mar 18. doi: [10.1093/aje/kw254](https://doi.org/10.1093/aje/kw254)

PMCID: PMC4832051

PMID: [26994063](https://pubmed.ncbi.nlm.nih.gov/26994063/)

Using Big Data to Emulate a Target Trial When a Randomized Trial Is Not Available

[Miguel A. Hernán*](#) and [James M. Robins](#)

Step 1:

Design de l'essai cible

« comme si l'essai contrôlé randomiser pouvait être fait... »

1. Eligibility criteria
2. Treatment strategies
3. Assignment procedures
4. Follow-up period
5. Outcome
6. Causal contrasts of interest
7. Analysis plan

Step 2:

**Compilation des données
baseline** à partir de données
observationnelles

Step 3:

Simulation des outcomes à
partir d'une base de données
observationnelles

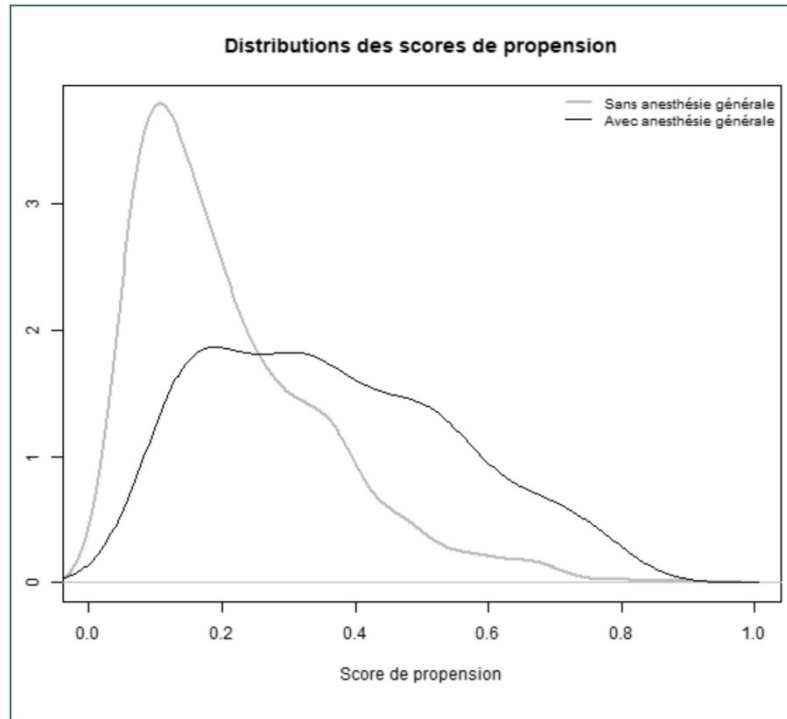
Comme toute méthode d'inférence causale, 4 hypothèses

1. Absence de facteurs de confusion non mesurés
2. Positivité
 - Pas de biais de sélection
3. Consistance
 - Définition claire des niveaux de traitement
4. Modélisations correctes (robustness or double robustness)

Outils les plus courants :

- Score de propension
- G-computation

Le **score de propension** est la probabilité qu'un patient reçoive un traitement donné en fonction de ses caractéristiques initiales.

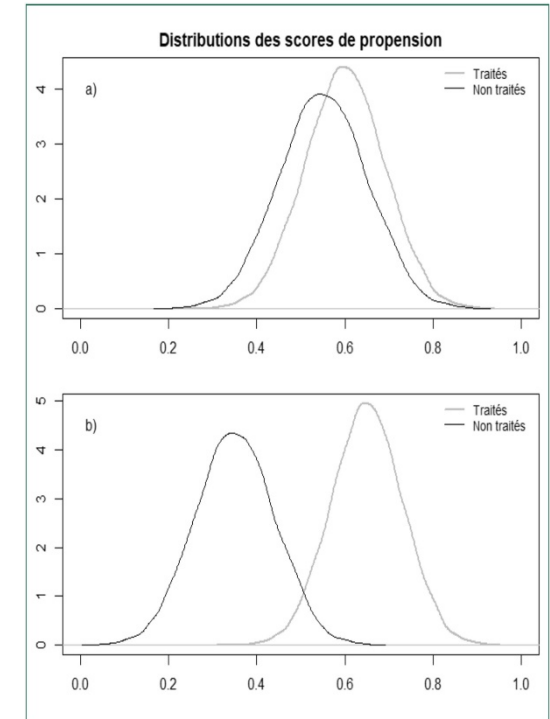


Identification de patients de chaque groupe avec des SP proches

- **Matching**

Corriger les estimations en pondérant par le SP

- **IPW**



Merci de votre attention...



"Your doctor will be here in a minute, I'm a placebo."