

INTRODUCTION À L'ÉPIDÉMIOLOGIE

Jérémie Botton

Épidémiologiste, [Epi-Phare ANSM-CNAM](#)

jeremie.botton@universite-paris-saclay.fr

L'objectif de ce cours

- Est:
 - De définir « l'épidémiologie »
 - D'avoir un aperçu des outils sur lesquels la discipline repose
 - De connaître les intérêts et limites des études épidémiologiques
 - introduction à la lecture critique d'article
- Il n'est pas:
 - De présenter les fondements statistiques sur lesquels cette discipline repose
 - (important de savoir qu'ils existent)

Plan

- Principes et généralités
 - Définitions
 - Pourquoi utiliser les méthodes statistiques
 - Bref historique de la discipline
- Données et principaux types d'études
 - Ecologiques
 - Transversales
 - Cas-témoins
 - Cohortes
 - (Etudes de séries de cas autocontrôlées)
- Paramètres calculés dans les enquêtes, biais possibles
 - Indicateurs de risque
 - Principaux types de biais
 - Interprétation des résultats
- Exemple

DÉFINITION DE L'ÉPIDÉMIOLOGIE

Les épidémiologistes recherchent pourquoi certains groupes de personnes sont plus à même que d'autres de devenir malades, d'être blessés ou de décéder jeunes.

Cela nécessite de réfléchir à la manière dont les mécanismes biologiques, les processus sociaux et les comportements mènent les personnes à la maladie ou à la blessure.

Le but de l'épidémiologie est de guider la Santé Publique et les politiques de soin afin d'enrayer ces mécanismes et améliorer la santé des populations et les événements de santé des patients.

Les épidémiologistes travaillent sur pratiquement tous les événements de santé imaginables, de la mortalité infantile à la maladie d'Alzheimer. Nous recherchons pratiquement toutes les expositions auxquelles vous pouvez penser, de la pollution de l'air aux polymorphismes génétiques, car il existe une multitude de facteurs modifiables qui influencent la santé.

Définition de l'épidémiologie

- Étude de la distribution et des déterminants d'états de santé ou d'événements liés à la santé dans des **populations** données, et application de cette étude à la lutte contre les problèmes de santé
- Remarques:
 - Pas seulement les maladies infectieuses;
 - Pas seulement les maladies très fréquentes (« épidémies »).
 - Toutes les pathologies humaines: maladies chroniques, génétiques, accidents, maladies dues aux comportements, etc.
 - Tous les facteurs susceptibles d'expliquer leur survenue (hypothèses a priori): environnement, alimentation, traitements, facteurs sociaux, gènes, etc.
 - Aide à la prise de décision en Santé Publique

Définition de l'épidémiologie

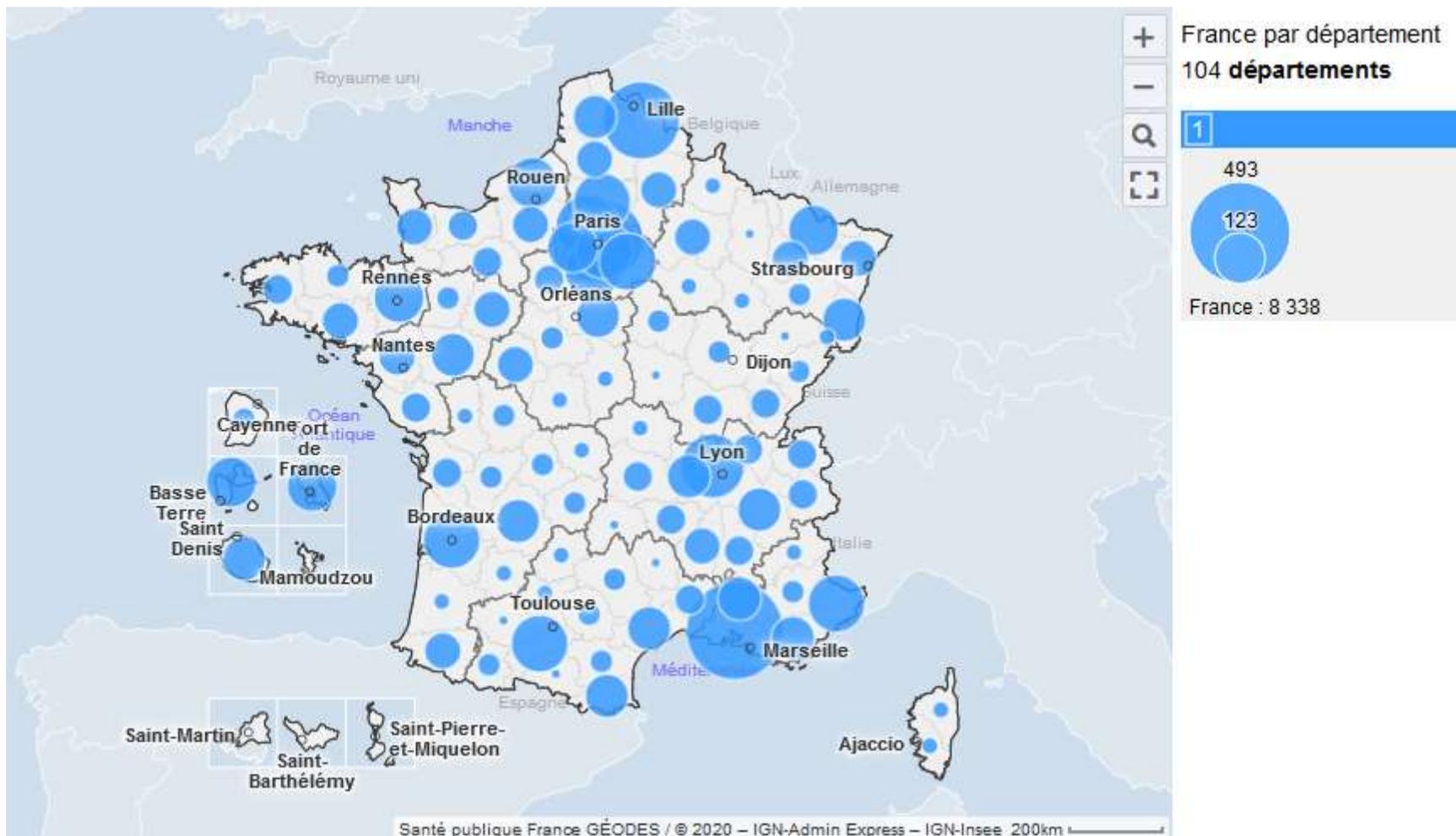
- Étude de la **distribution** et des **déterminants** d'états de santé ou d'événements liés à la santé dans des populations données, et application de cette étude à la lutte contre les problèmes de santé (Last, 2001)

Epidémiologie descriptive

Epidémiologie analytique
(ou étiologique)

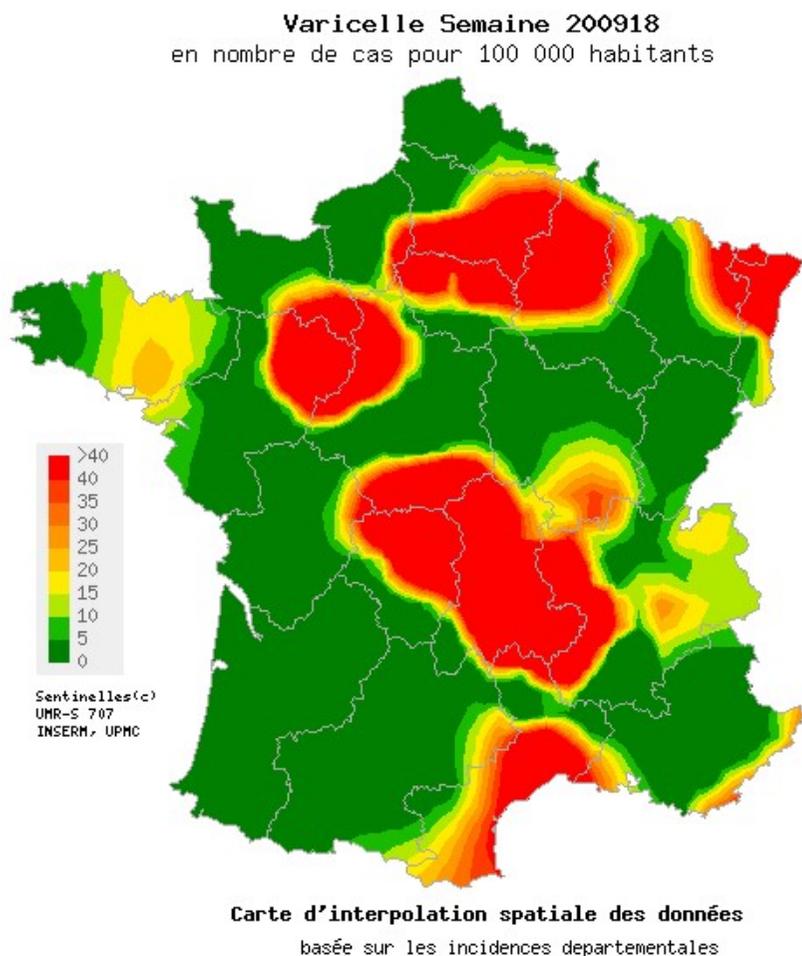
Distribution (ex: COVID-19)

Nombre de personnes hospitalisées avec diagnostic COVID-19 - hommes et femmes à la date du 22/11/2021

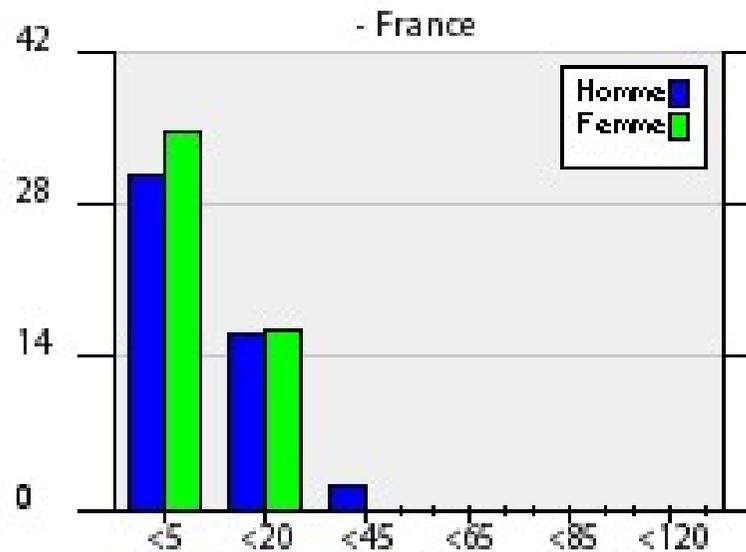
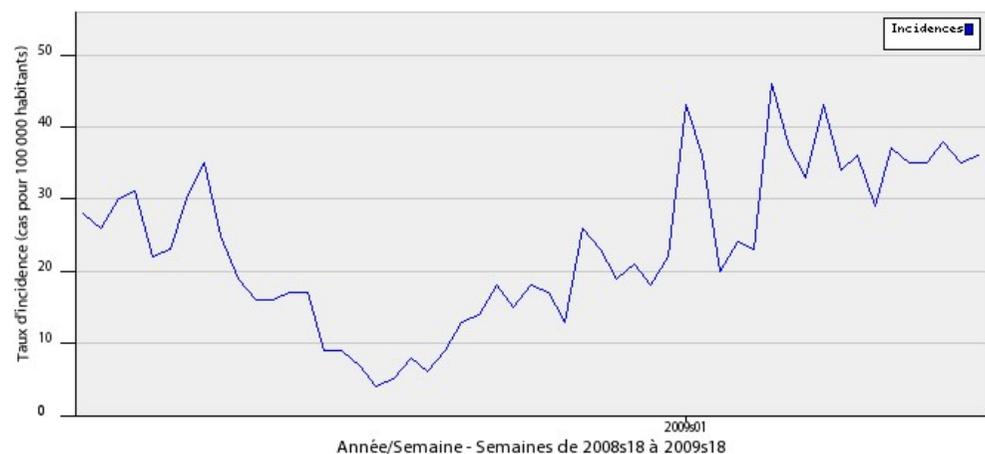


<https://geodes.santepubliquefrance.fr/#view=map2&c=indicator>

Distribution (ex: varicelle, réseau sentinelle)



Réseau Sentinelles, Varicelle, France entière



Déterminants de la santé

- Liés aux individus: âge, sexe, groupe ethnique, caractéristiques génétiques, sociales, biologiques, physiques, etc.
- Liés au lieu: pays, climat, lieu de résidence, région géographique, éducation, travail, espaces verts
- Liés au temps: variations annuelles, saisonnières, apparitions brutales

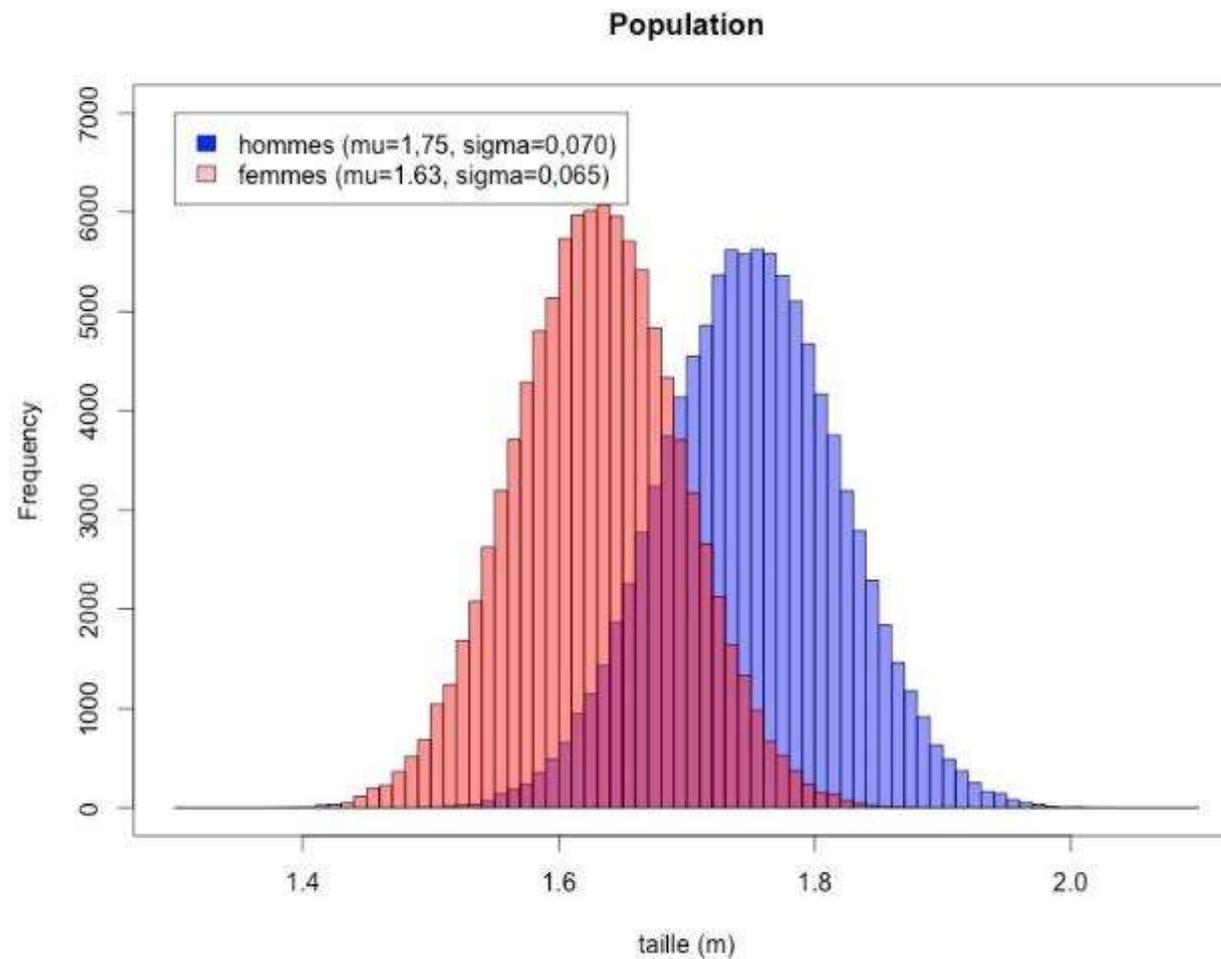
Épidémiologie analytique: exemples

« Exposition »	« Etat de santé »
◆ Tabagisme	Cancer du poumon
Amiante	Mésothéliome
Cholestérol	Maladies cardiovasculaires
Durée du sommeil	Diabète
◆ Vaccin hépatite B	Sclérose en plaques
Allaitement artificiel	Développement cognitif altéré
Croissance précoce réduite	Maladies cardiovasculaire
Anxiolytiques	Accidents de la route
Traitement de la ménopause	Thrombose
Papillomavirus	Cancer du col de l'utérus
◆ Benfluorex	Valvulopathies
Thalidomide (sédatif, anti-nauséeux)	Phocomélie
Plomb	Saturnisme
◆ Perturbateurs endocriniens	Obésité

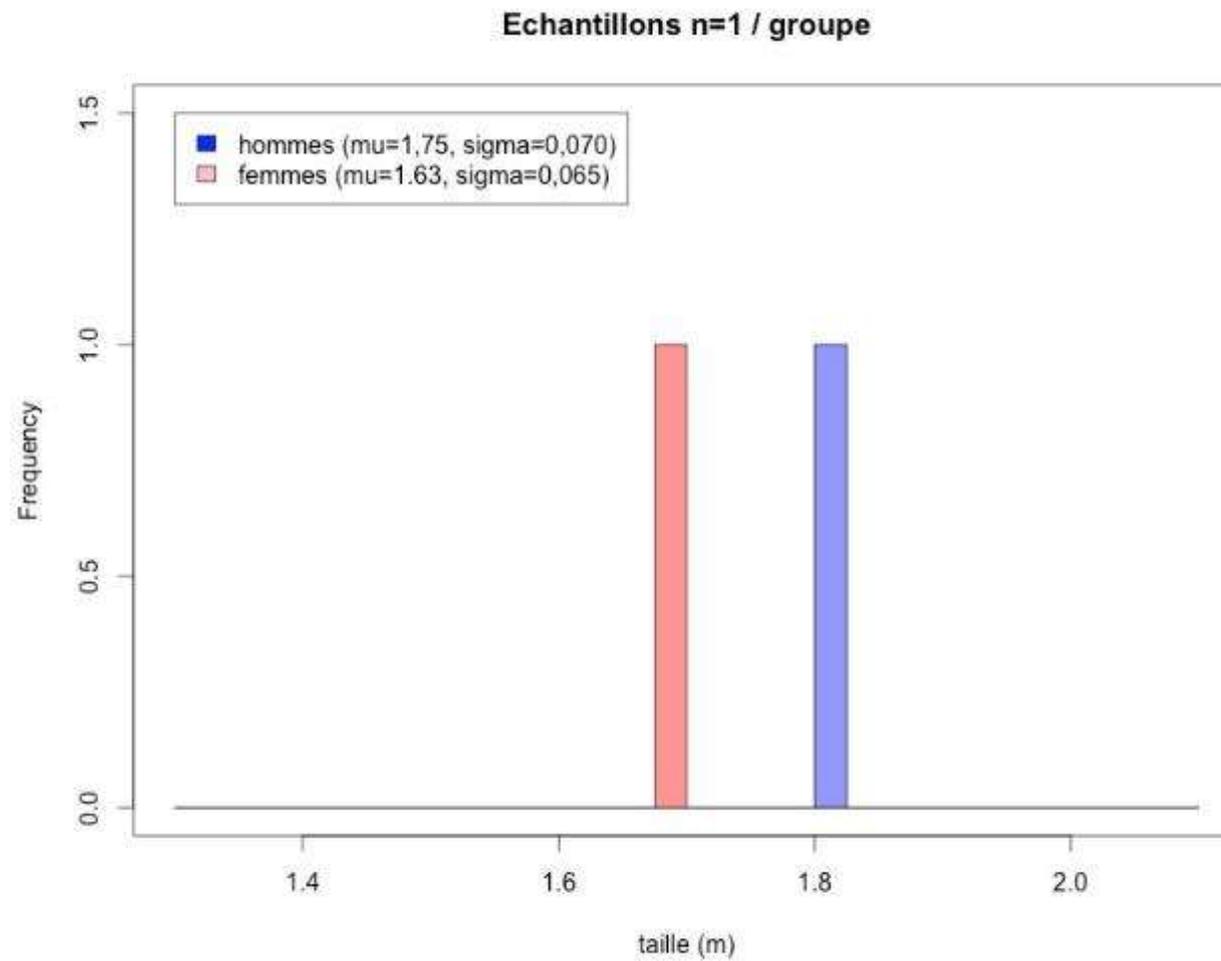
Risque
de...?

POURQUOI DES OUTILS STATISTIQUES ?

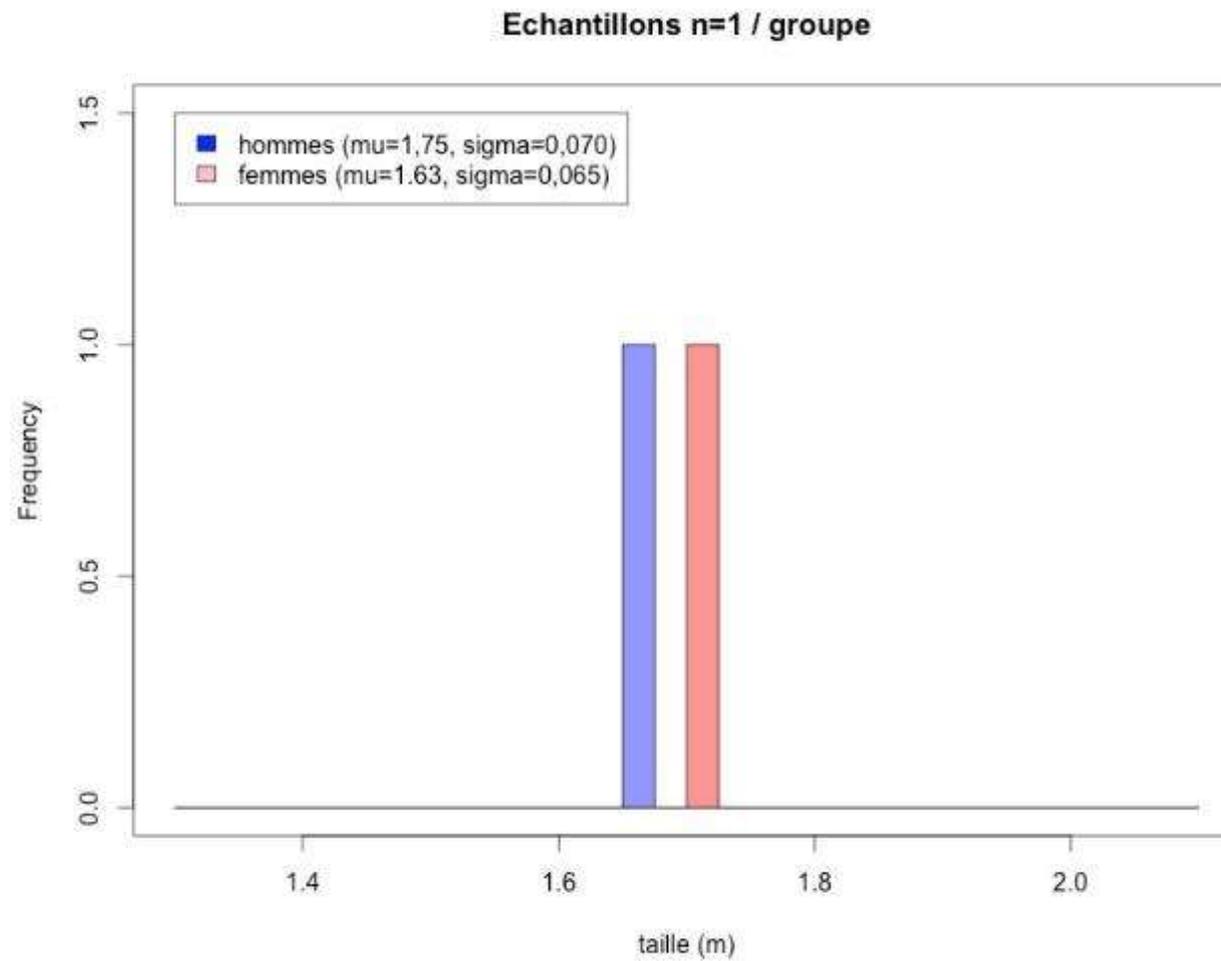
Différence de taille entre hommes et femmes?



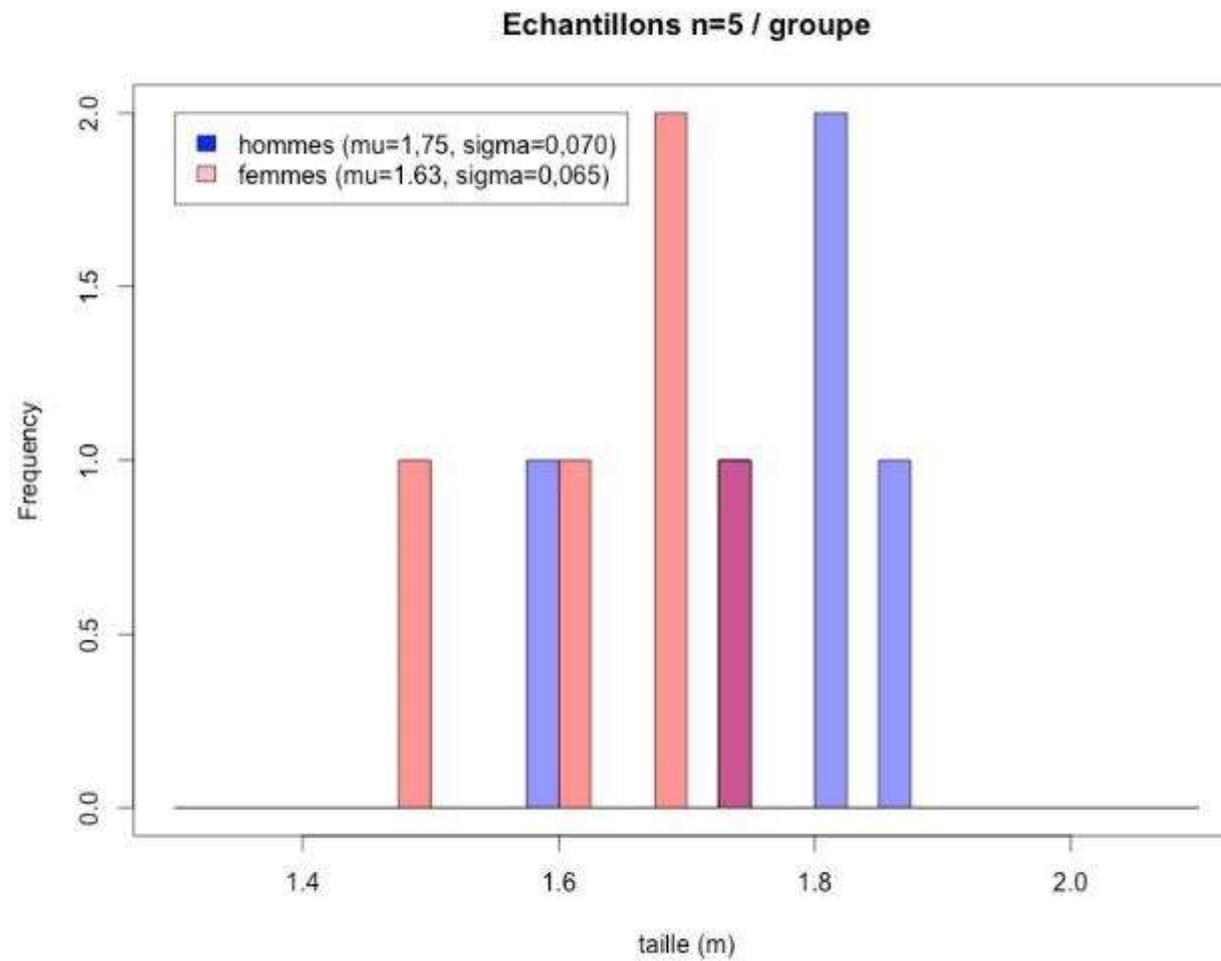
Différence de taille entre hommes et femmes?



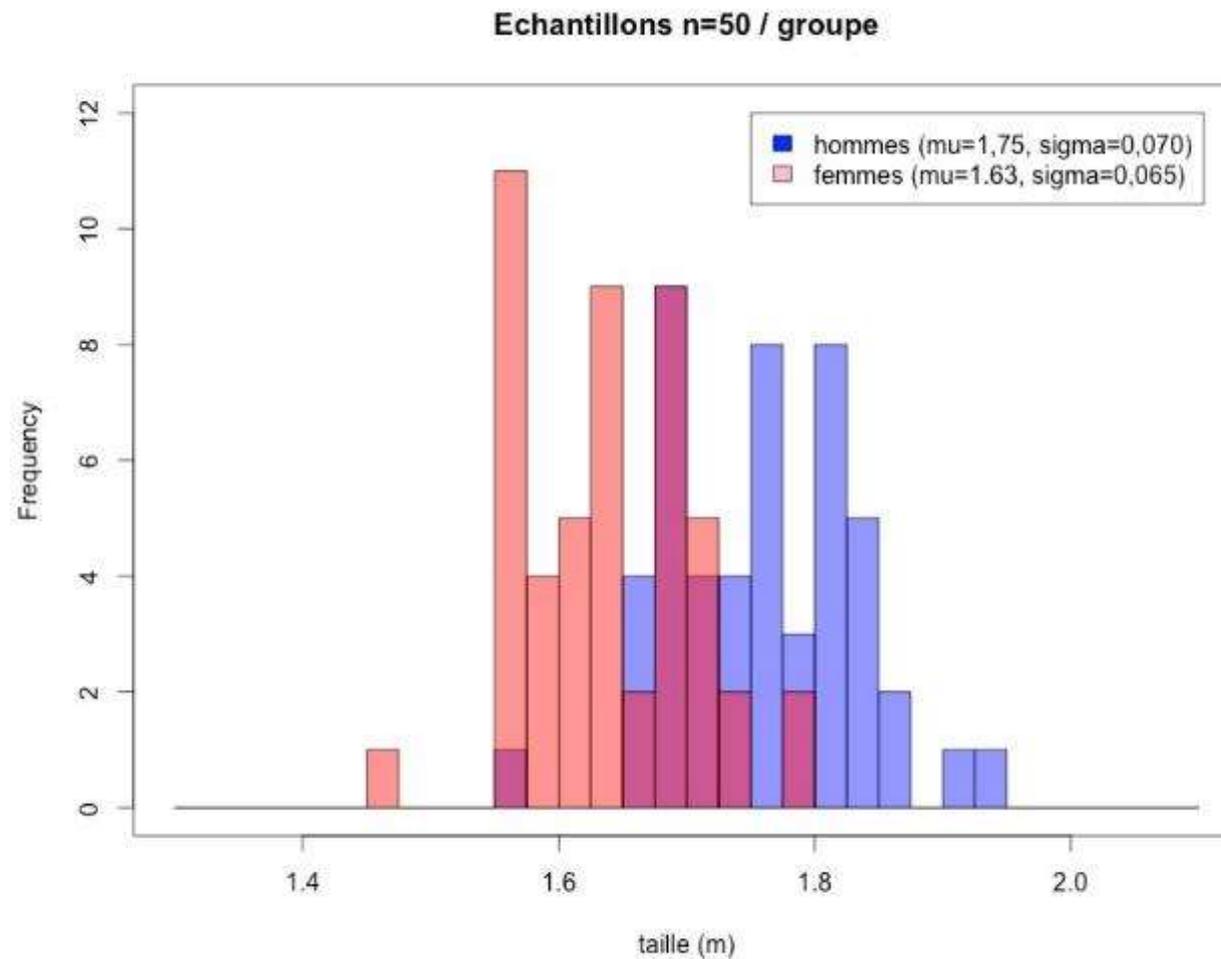
Différence de taille entre hommes et femmes?



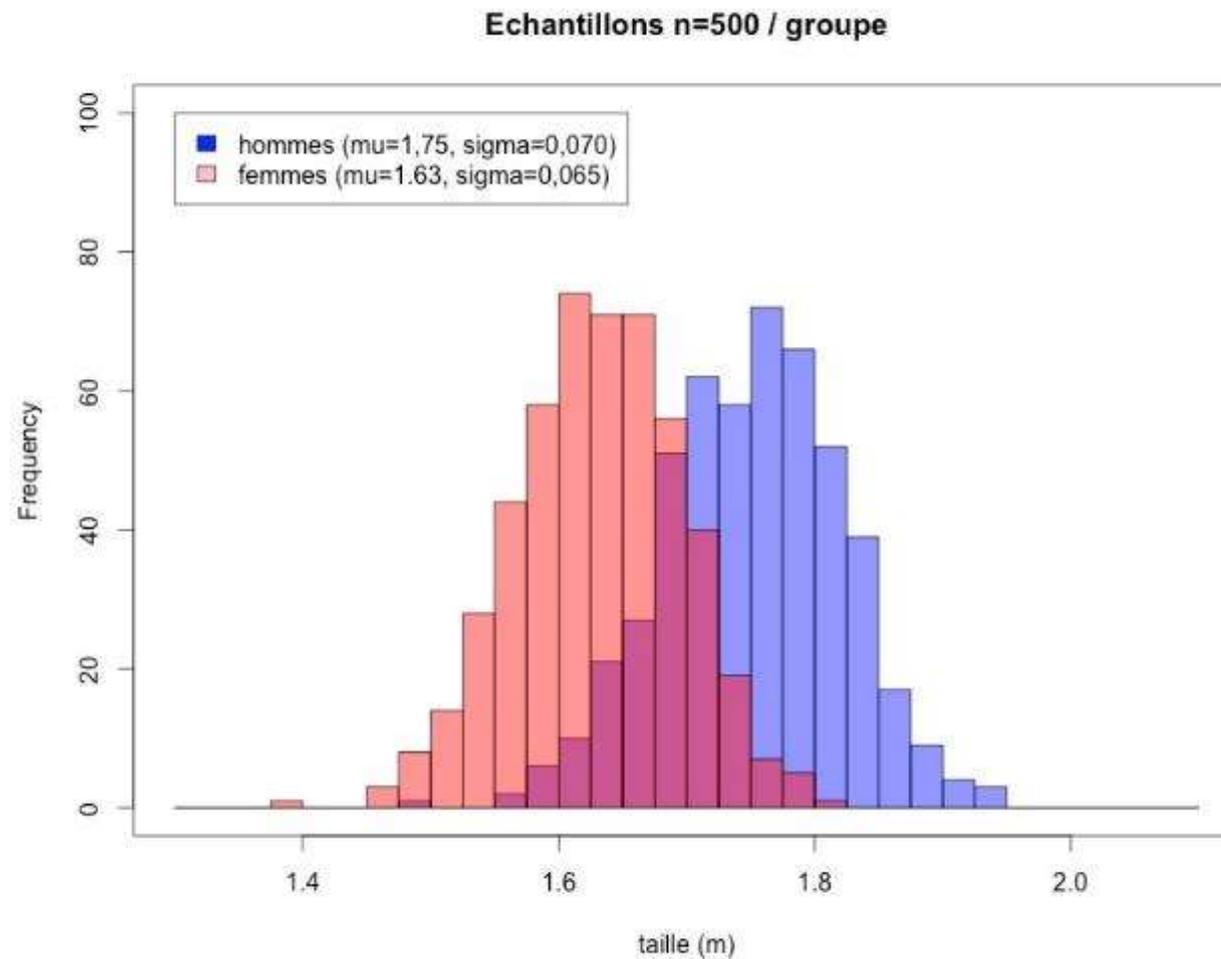
Différence de taille entre hommes et femmes?



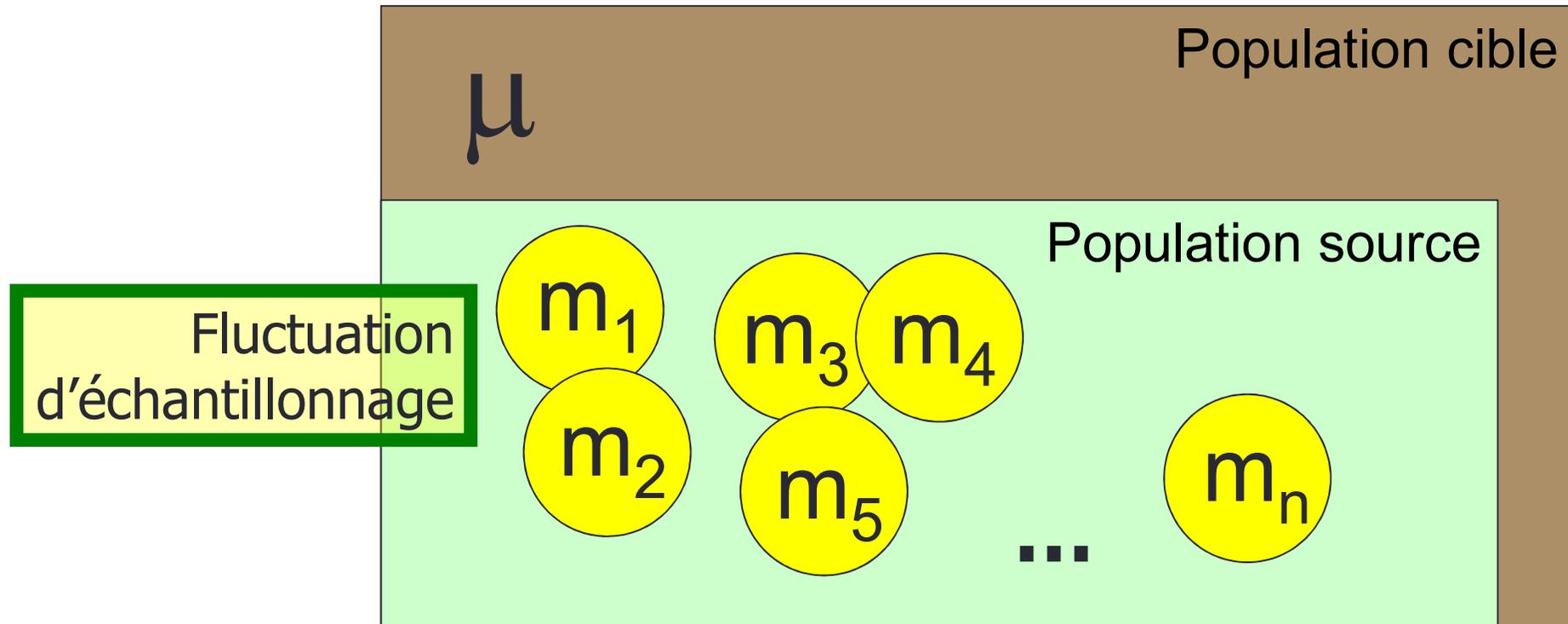
Différence de taille entre hommes et femmes?



Différence de taille entre hommes et femmes?



Fluctuation d'échantillonnage



D'autant plus faible que l'échantillon est grand
(l'incertitude de l'estimation diminue)

Etudier les facteurs influençant une variable

- **Variabilité** dans la population (« variable » de santé)
- Tenter de l'expliquer par certains « facteurs » (exposition)
- Exemple:
 - Échantillon représentatif des 18-25 ans
 - Deux groupes: surpoids (IMC > 25 kg/m²) / non
 - Moyenne de cholestérol dans ces 2 groupes
 - Question: la moyenne est-elle différente entre les 2 groupes?
 - Question concomitante importante: **les 2 groupes sont-ils comparables** (âge, sexe, statut socio-économique, etc.)?
- Une différence sera jugée « significative » si elle dépasse « d'un certain degré » la variabilité liée au hasard
- Plus l'effectif est important, plus grande est la précision sur ce que l'on cherche à estimer (moyenne, pourcentage, indicateur de risque): on réduit l'incertitude.

Tests d'hypothèse

Rappel pour la comparaison de 2 moyennes

- Condition: les 2 variances sont identiques
- H_0 : « égalité entre les 2 moyennes (dans la population) »

- Statistique:

$$\frac{m_1 - m_2}{\sqrt{s^2 \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}} \sim t_{n_1+n_2-2}$$

Rappel:

s^2 : variance de la mesure

s^2/n : variance de la moyenne

- Principe: si H_0 est vraie, l'expérience devrait donner des valeurs de moyennes pas très différentes
- Si ce n'est pas le cas, alors on va rejeter H_0 avec une prise de risque α (a priori)
- **Le plus souvent, le test est interprétable si les 2 groupes sont comparables**

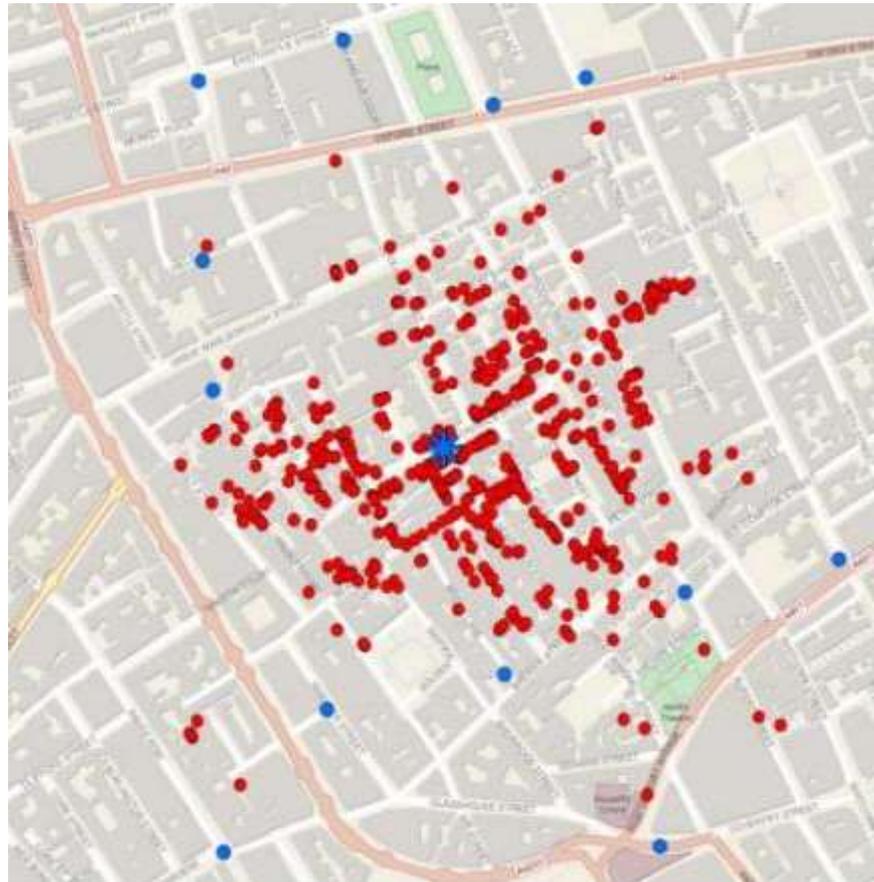
HISTOIRE

Histoire: de Hippocrate au XX^e siècle

Étymologiquement: épidémie = « qui tombe sur le peuple »

- Etiologies selon Hippocrate (V av. JC) et Galien (III av JC):
 - Hippocrate: conditions atmosphériques, nutrition et style de vie
 - Galien: atmosphérique, susceptibilité et style de vie
- Paracelse (1534): 1^{er} traité sur les maladies professionnelles
- Essai clinique de James Lind (1747): scorbut et vitamine C
 - 6 groupes: cidre, acide sulfurique, vinaigre, herbes/épices, eau de mer, oranges/citrons
- Daniel Bernoulli (1766): modélisation mathématique de la propagation des maladies, impact de la variole
- William Farr (1840): collecte systématique de données (registre de mortalité, registres de naissance)
- Cause du choléra (John Snow, Londres 1853): origine hydrique (20 ans avant le microscope)

Carte de John Snow



Locations of water pumps and cholera deaths from John Snow's map, from <http://donboyes.com/2011/10/14/john-snow-and-serendipity>

Epidémiologie moderne (1930-1950)

- **Apparition du terme « épidémiologie » en 1802**, réservé jusqu'au milieu du XX^e siècle aux maladies infectieuses
- Études de Doll et Hill: relation entre cancers broncho-pulmonaires et consommation de tabac
 - Cas-témoins appariée
 - Cohorte de 20 000 médecins anglais
- Cohorte Framingham (1949, suivi ~ 50 ans): explorer les déterminants des maladies cardiovasculaires (5127 sujets)
- Grandes cohortes généralistes récentes:
 - Etude longitudinale française depuis l'enfance: <https://www.elfe-france.fr>
 - Cohorte de 200 000 adultes: <http://www.constances.fr>
- Données du Système National des Données de Santé (SNDS)

DONNÉES

Support d'étude: la population

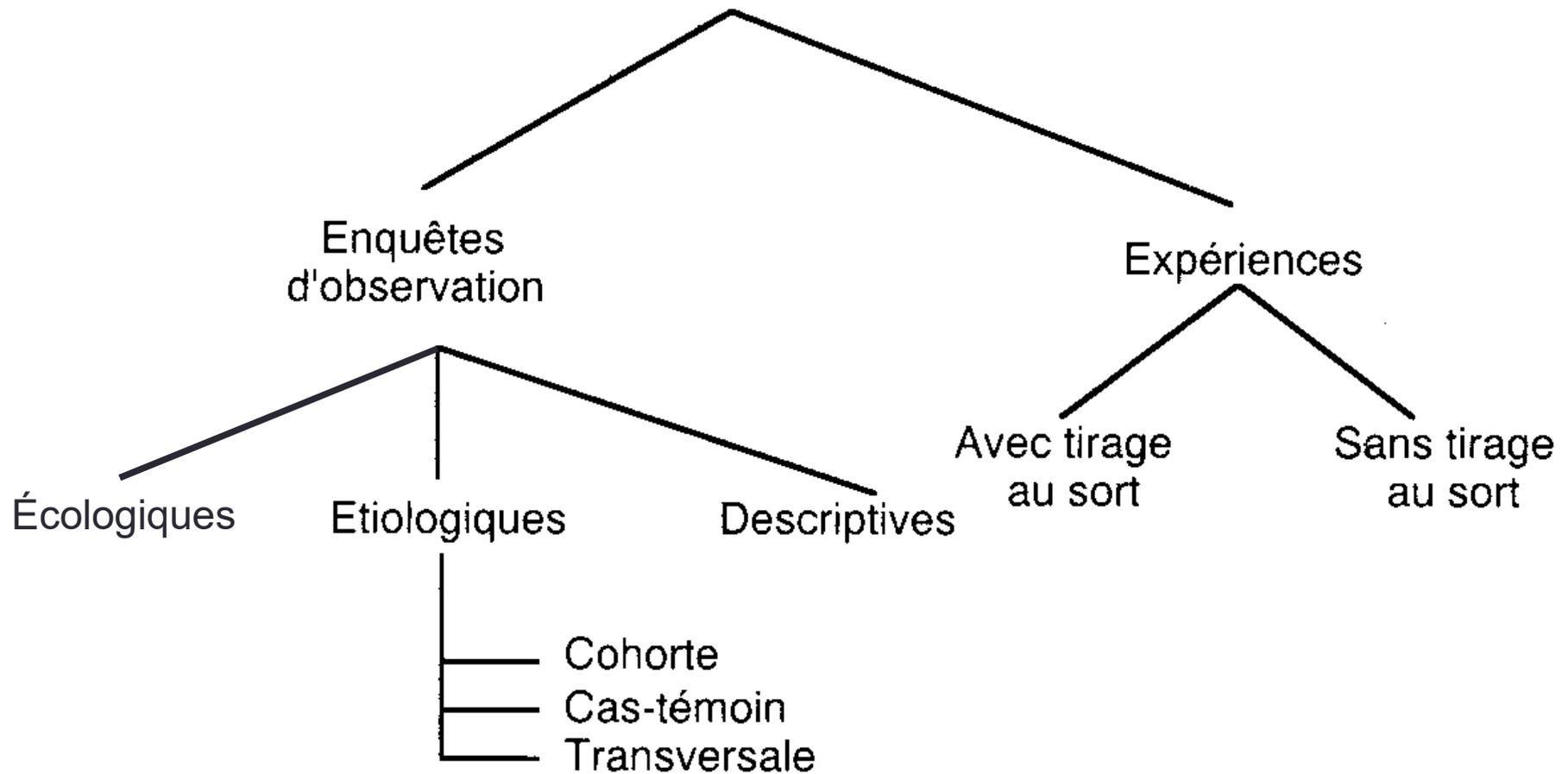
- Souvent difficile/impossible d'accéder à toute la population → sondage → échantillon de la population
- Statistique: à partir de l'échantillon en déduire des propriétés de la population
- On ne peut rien conclure à partir de l'observation d'un individu
- Remarques sur les essais contrôlés randomisés:
 - Population de malades, très sélectionnés, durée courte (→ pas pour les effets indésirables à long terme),
 - Importance des études de pharmacoépidémiologie

Sources de données

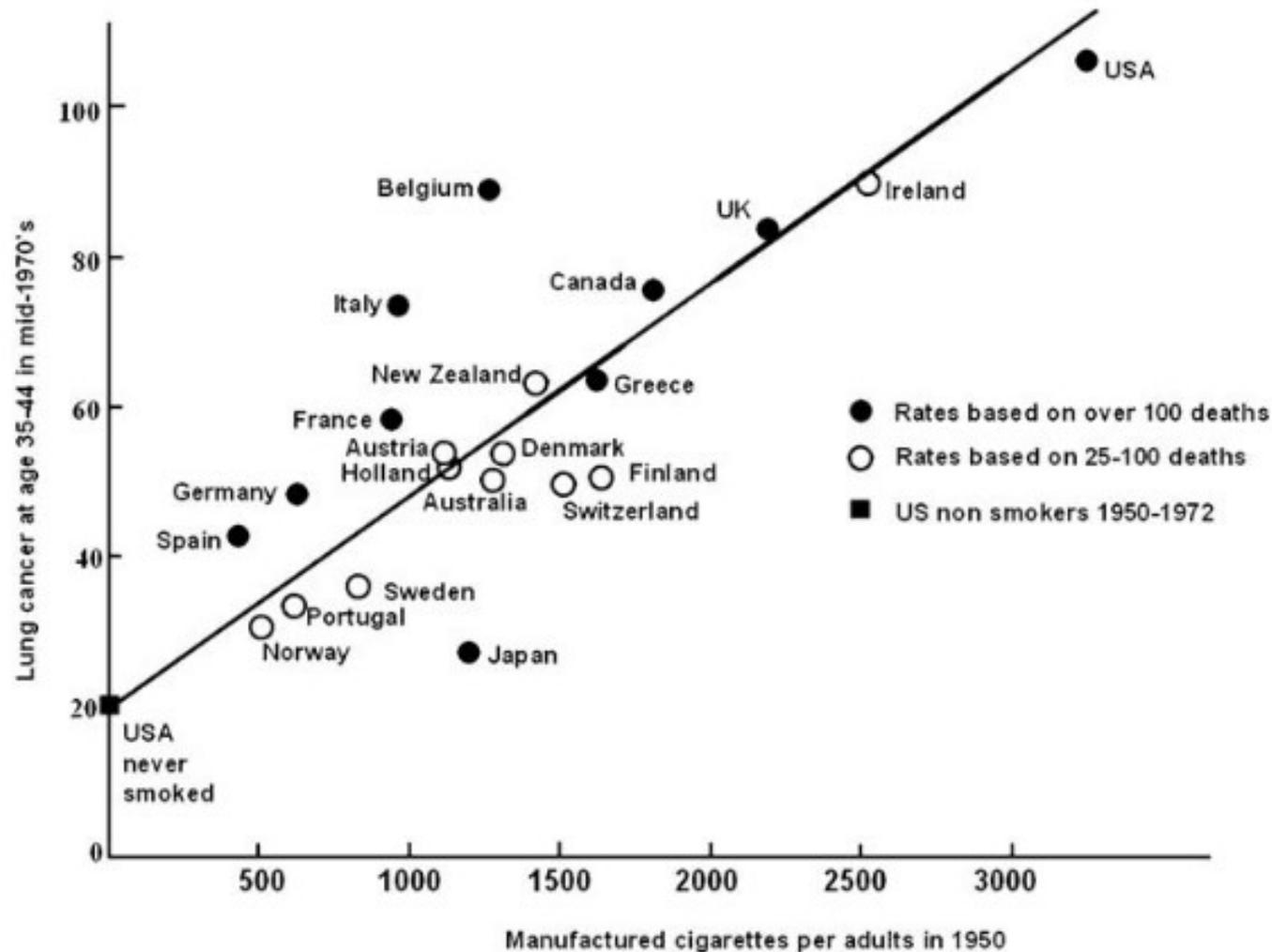
- Statistiques de mortalité et causes de décès:
 - Médecin → Mairie → ARS et Insee → Inserm (CépiDC: centre d'épidémiologie sur les causes médicales de décès)
 - Pas adapté pour l'étude de morbidité de toutes les pathologies
 - Données disponibles dans le SNDS
 - Registres de morbidité:
 - tous les cas d'1 pathologie dans 1 région donnée (ex: cancers, maladies cardiovasculaires, malformations congénitales)
 - Informations médicales individuelles:
 - médecine du travail, assurance maladie, dossier médical, ...
 - Système national des données de santé, Health Data Hub
 - Réseaux sociaux
 - Protocoles spécifiques
-
- NB. L'ANSM souhaite développer la « culture de la donnée »

ENQUÊTES

Principaux types d'enquêtes

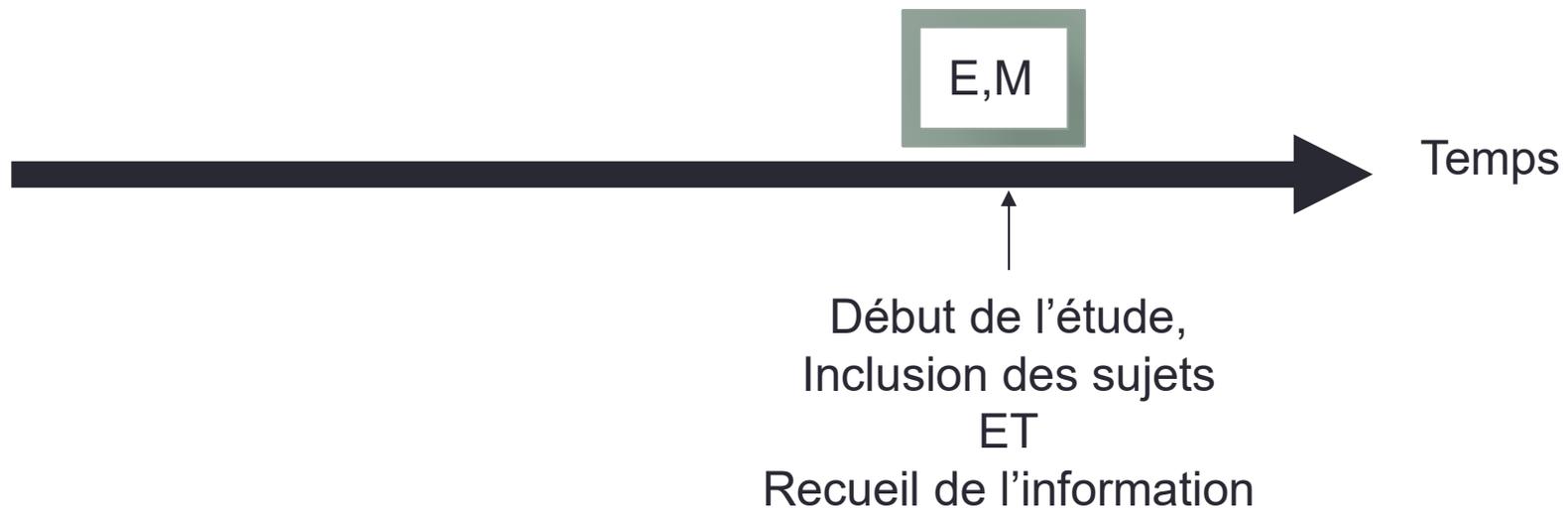


Enquêtes écologiques (données agrégées)



Enquêtes transversales

Recueil SIMULTANÉ



	E+	E-
M+	a	b
M-	c	d

a-d: effectifs

Enquêtes transversales

- **Principe:** recueil de l'état de santé et des variables d'exposition, à un moment donné, dans une population définie
- Avantages:
 - Prévalences de la / des maladie(s), de la / des exposition(s) si échantillon représentatif
 - Rapide (pas de suivi)
- Inconvénients:
 - Pas pour maladies ou expositions rares ou courtes
 - Pas de chronologie des événements (cause ou conséquence?)

Enquêtes cas-témoins



Recueil de E
RÉTROSPECTIF

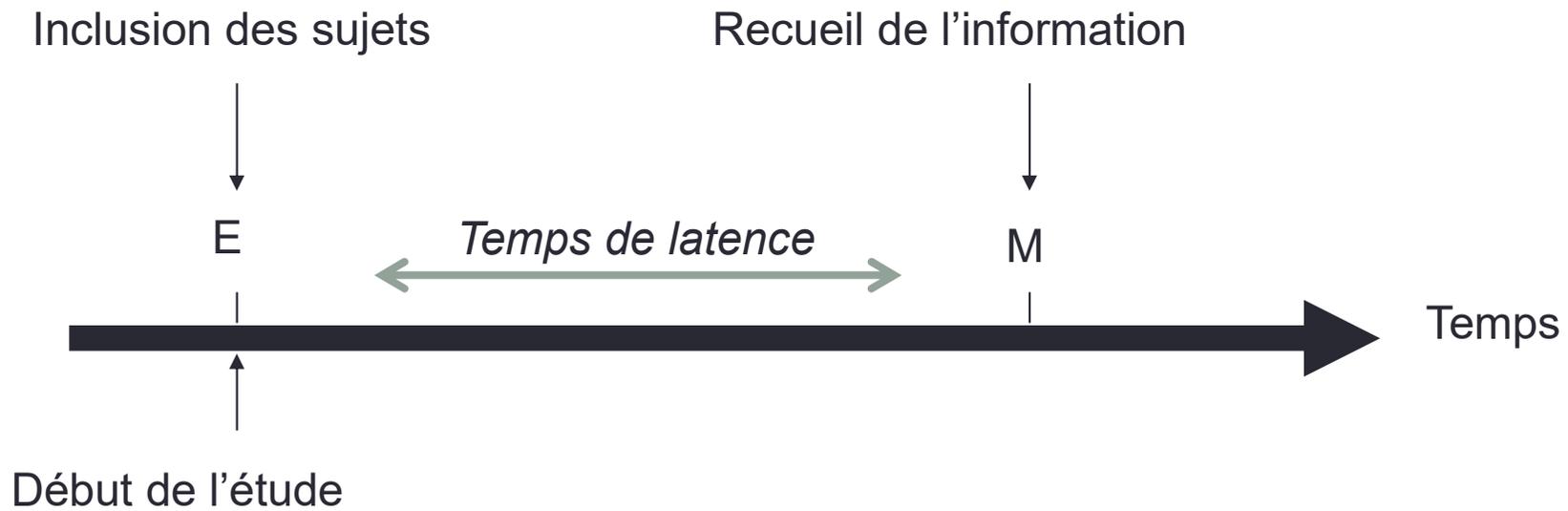
	E+	E-
M+	a	b
M-	c	d

a-d: effectifs

Enquêtes cas-témoins

- **Principe:** un groupe de personnes atteintes de la maladie (cas) et un groupe de personnes non atteintes (témoins); comparaison de la fréquence d'exposition entre les deux (rétrospectif).
- Individus généralement appariés: pour un cas, sélection d'un ou plusieurs témoins partageant des caractéristiques du cas (age, sexe...)
- Avantages:
 - Rapide
 - Adapté aux maladies rares
(sur-représentation des malades / population)
 - Etude de plusieurs expositions
- Inconvénients:
 - Pas adapté aux expositions rares
 - Choix des témoins difficile: le plus proche possible des cas
 - Biais de mémoire: les cas exposés se souviennent mieux de leur exposition (= biais de classement, défini plus loin)

Cohortes prospectives



Suivi en temps réel

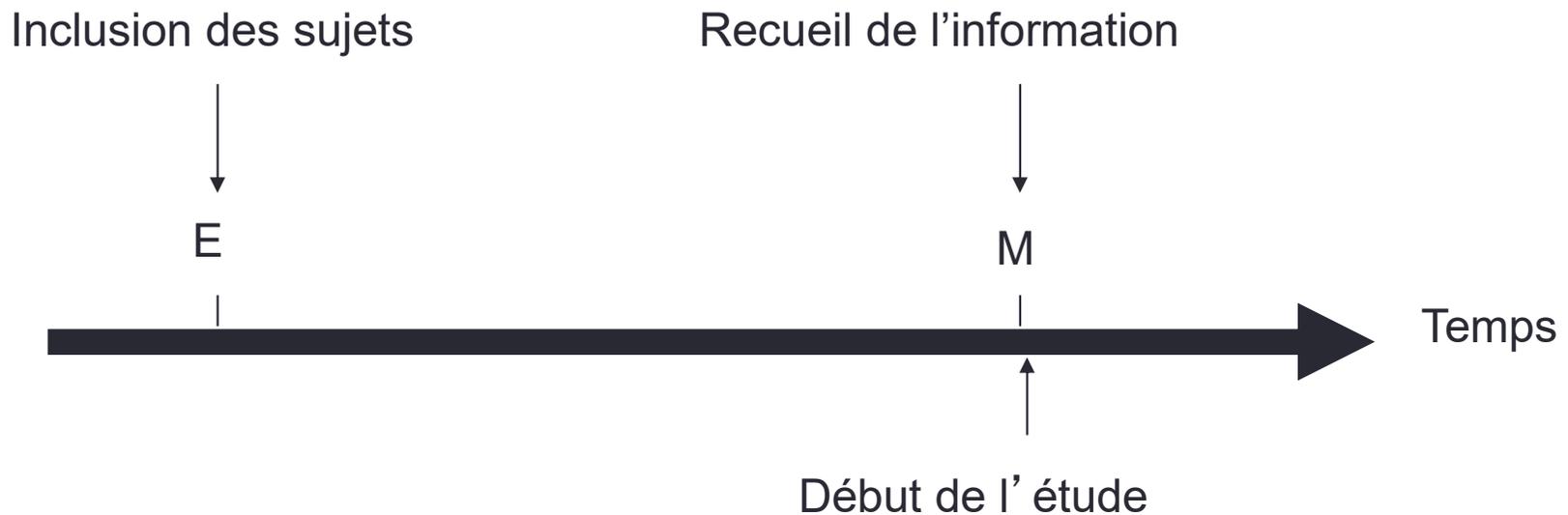
	E+	E-
M+	a	b
M-	c	d

a-d: effectifs

Cohortes prospectives

- **Principe:** Suivi d'une population non atteinte de la maladie sur une (\pm longue) période et comparaison de l'incidence (vitesse d'apparition) de la maladie selon l'exposition.
- **Avantages:**
 - Adapté aux expositions rares
(sur-représentation possible / population: études exposés / non exposés)
 - Groupe de référence plus facile à constituer
 - Etude de plusieurs maladies
- **Inconvénients:**
 - Pas adapté aux maladies rares
 - Long et sujets perdus de vue au cours du temps
 - Coûteux

Cohortes rétrospectives / historiques



	E+	E-
M+	a	b
M-	c	d

a-d: effectifs

**On retrace le suivi
« comme si » en temps réel**

INDICATEURS DE RISQUE

Indices descriptifs et étiologiques

- Indices descriptifs:
 - Prévalence (p),
 - Taux d'incidence (TI),
 - Moyenne (m),
 - etc.
- Indices étiologiques:
 - Risque Relatifs (RR),
 - Odds Ratio (OR),
 - Pente d'une régression linéaire (β),
 - Hazard Ratio (HR),
 - etc.

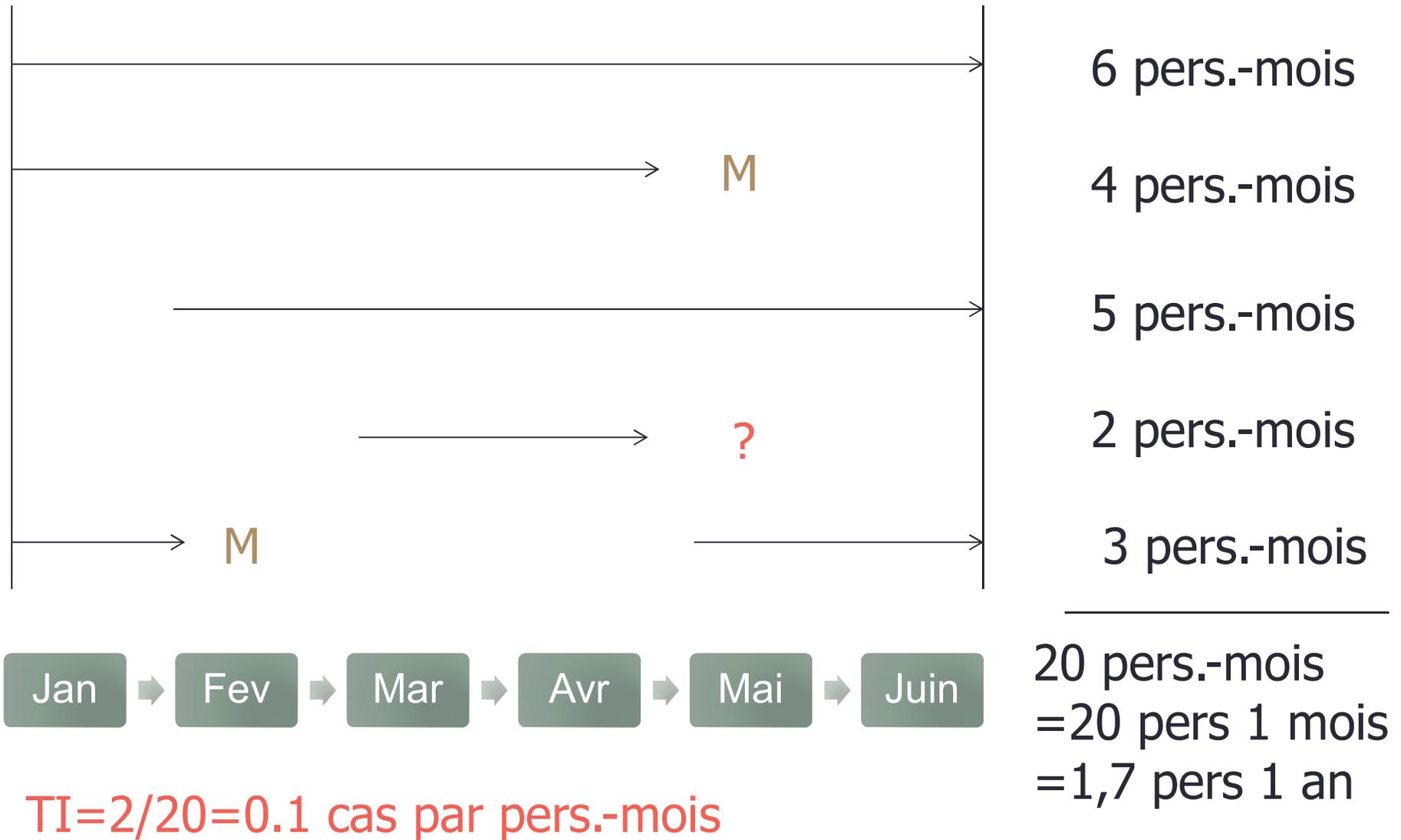
Taux d'incidence et prévalence

- Prévalence: proportion de cas à un instant donné
(= comprise entre 0 et 1)
- Taux d'incidence: vitesse d'apparition de nouveaux cas dans une population sur une période donnée (\neq proportion):

$$TI = \frac{\text{Nombre de cas}}{\text{Nombre de personnes-temps}}$$

Personnes-temps: somme, sur tous les sujets de l'étude, des durées au cours desquelles les sujets sont susceptibles d'être enregistrés comme nouveau cas

Exemple de calcul de TI



Intervalle de confiance du taux d'incidence

- Taux d'incidence $TI = \frac{m}{PA}$

- Variance $\text{var}(TI) = \frac{m}{PA^2}$

- Intervalle de confiance à 95%

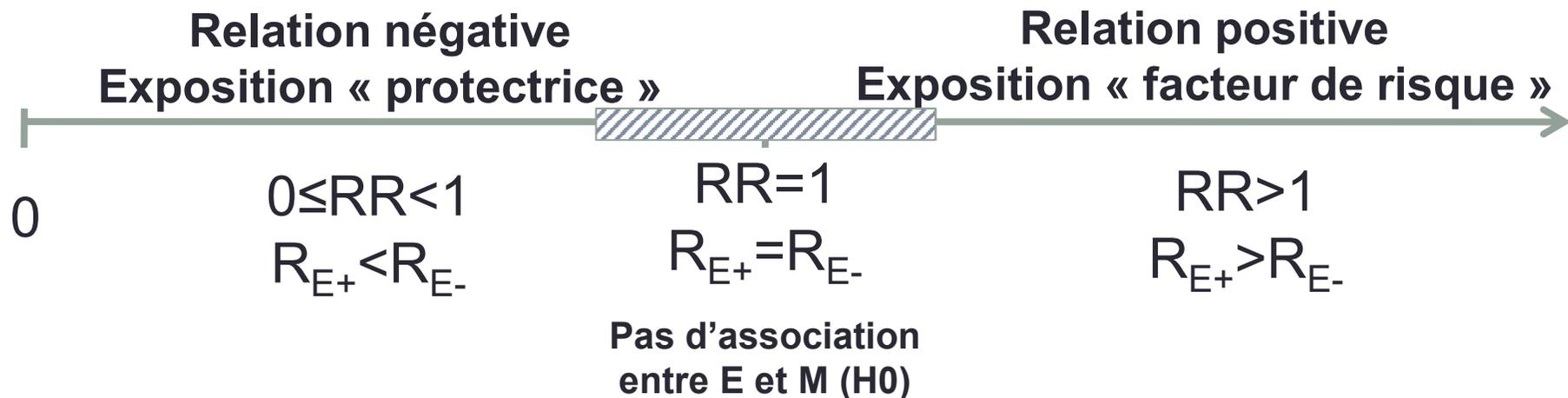
$$TI \pm 1,96 \cdot \sqrt{\text{var}(TI)}$$

Risque relatif

- Comparaison du risque de maladie chez les exposés par rapport aux non exposés:

$$RR = \frac{R_{E+}}{R_{E-}}$$

- Interprétation:



Interprétation du RR

- Si $RR=1,30$: Le risque chez les exposés est 1,30 fois plus grand que chez les non-exposés; les exposés ont un risque plus élevé de 30%.
 - On parle de relation positive: exposition et maladie évoluent dans le même sens
- Si $RR=0,90$: Le risque chez les exposés est 0,90 fois plus grand que chez les non-exposés; les exposés ont un risque plus faible de 10%.
 - On parle de relation négative: exposition et maladie évoluent dans un sens opposé
 - NB. $RR=0,10 \leftrightarrow$ « efficacité » de 90%

Calcul d'un risque relatif

Tableau des effectifs:

		Exposition	
		E+	E-
Maladie	M+	a	b
	M-	c	d

Exemple:

	Clopidogrel + aspirine (R=16,5%)	Placebo + aspirine (R=18,6%)
Infarctus	a=1035	b=1187
Pas d'infarctus	c=5224	d=5116

$$RR = \frac{R_{E+}}{R_{E-}} = \frac{\left(\frac{a}{a+c} \right)}{\left(\frac{b}{b+d} \right)}$$

- Impossible dans les enquêtes cas-témoins: M (donc R) sur-représenté

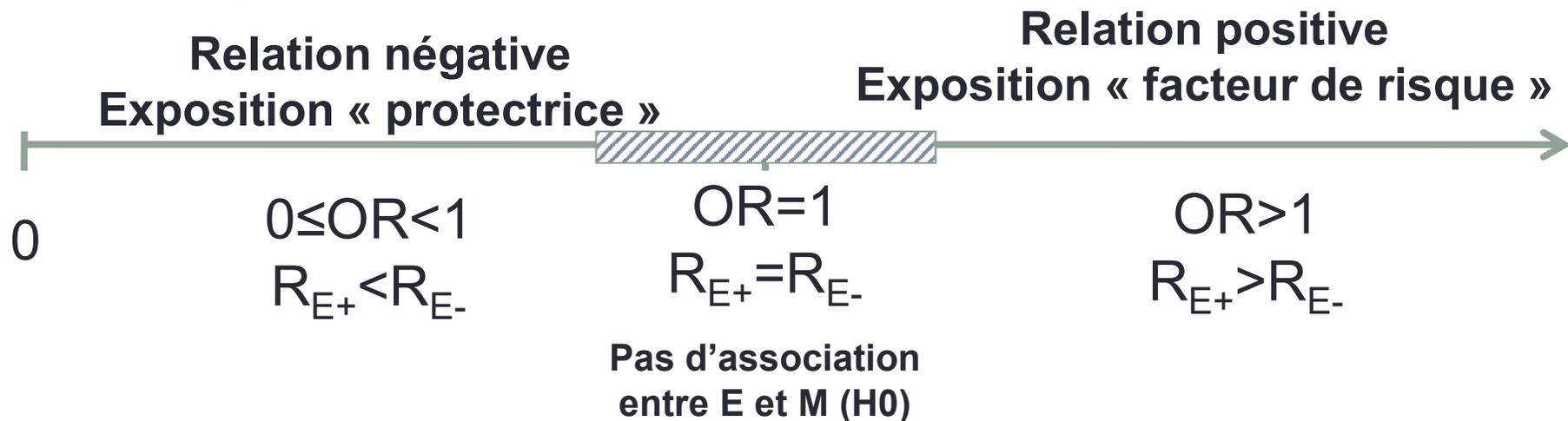
Odds ratio (OR)

- Comparaison des cotes (odds) de maladie chez les exposés par rapport aux non exposés:

	E+	E-
M+	a	b
M-	c	d

$$OR = \frac{\left(\frac{R_{E+}}{1 - R_{E+}} \right)}{\left(\frac{R_{E-}}{1 - R_{E-}} \right)} = \frac{ad}{bc}$$

- Interprétation:



Peut être estimé dans tous les types d'enquêtes

Comparaison de l'OR à 1: association significative?

	E+	E-
M+	a	b
M-	c	d

- H_0 : OR=1 *versus* H_1 : OR \neq 1
- Statistique:
 - Basée sur le logarithme népérien de l'OR :

$$z = \frac{\ln(OR)}{\sqrt{\text{var}[\ln(OR)]}} = \frac{\ln(OR)}{\sqrt{\frac{1}{a} + \frac{1}{b} + \frac{1}{c} + \frac{1}{d}}} \sim \mathcal{N}(0,1) \text{ sous } H_0$$

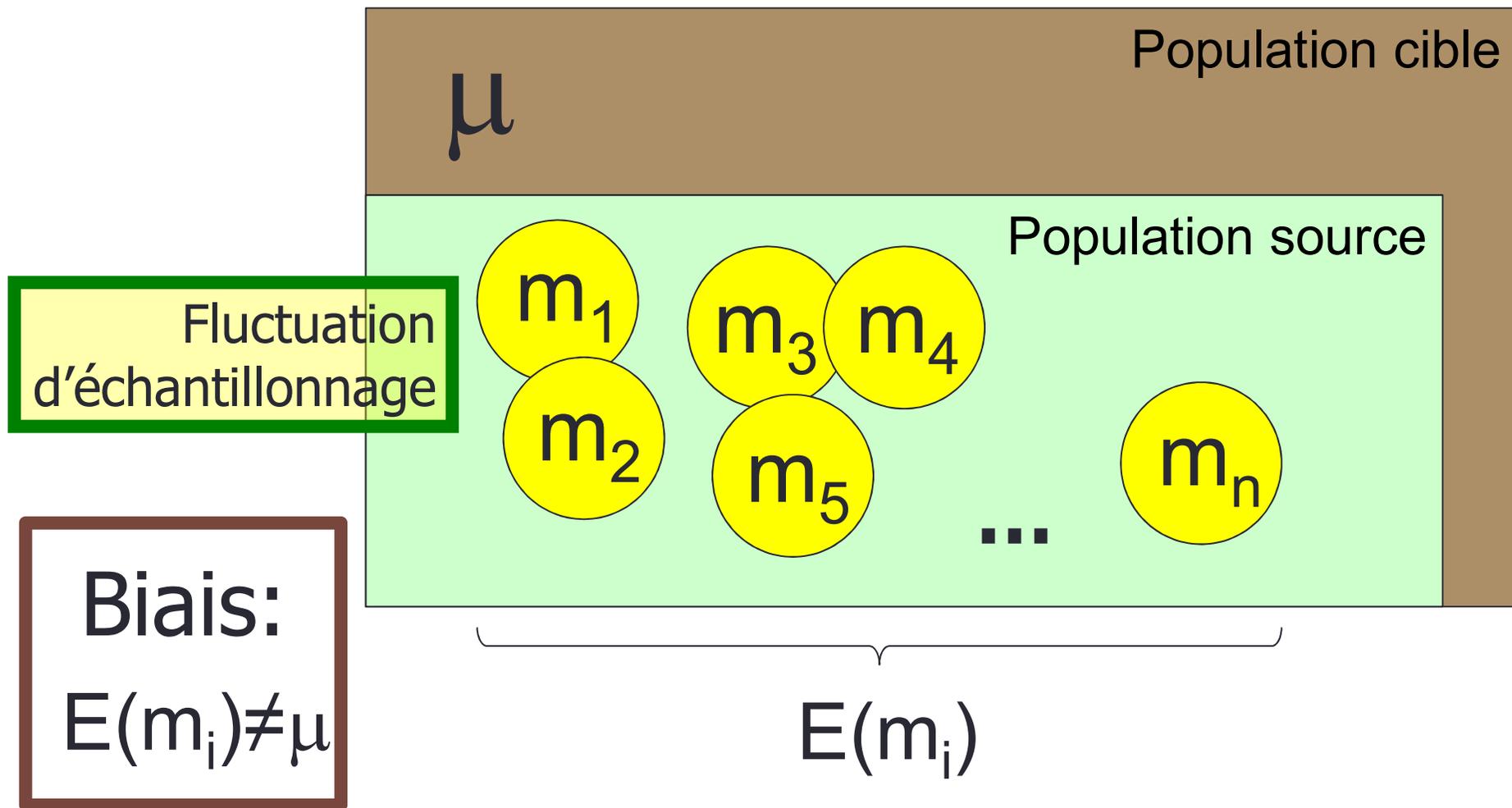
- Décision:
 - Si $|z_{\text{exp}}| > 1,96$, rejet H_0 : assoc. entre E et M au risque $\alpha=5\%$
 - Si $|z_{\text{exp}}| < 1,96$, non rejet de H_0 : pas d'assoc. significative

Interprétation de l'OR

- Pas la même interprétation que le RR
- S'en approche si deux conditions:
 - La prévalence de la maladie est rare ($\sim < 0,10$)
 - Le RR n'est pas très élevé ($< \sim 3$ environ)

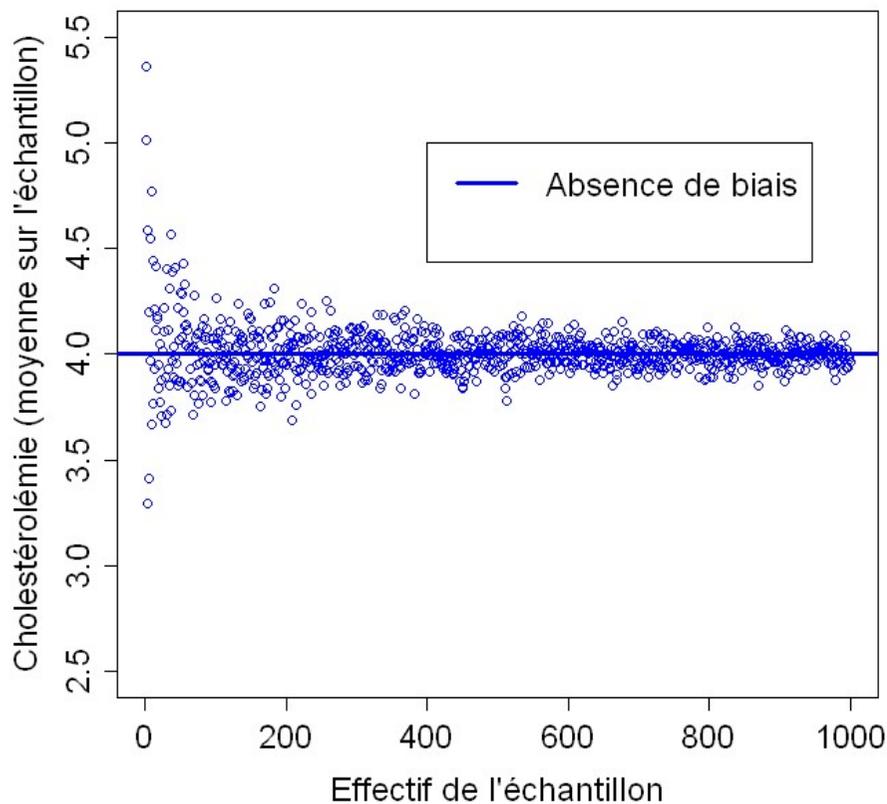
BIAIS

Fluctuation d'échantillonnage

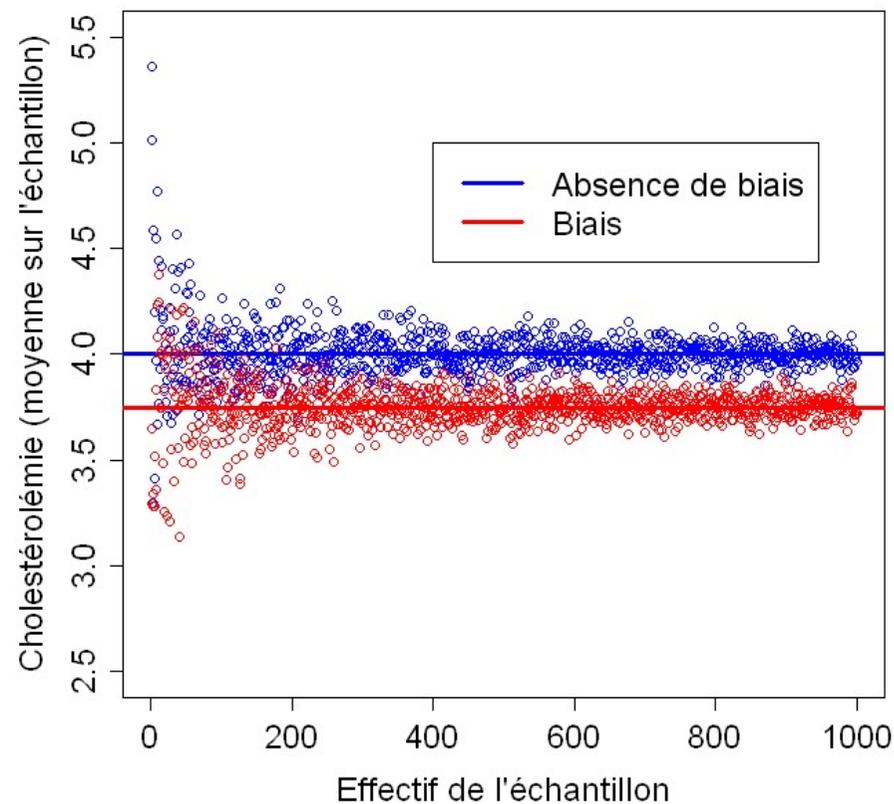


Fluctuation d'échantillonnage et biais

Fluctuation d'échantillonnage et biais

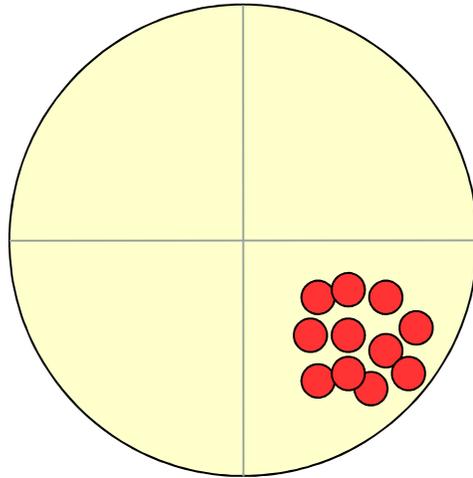


Fluctuation d'échantillonnage et biais

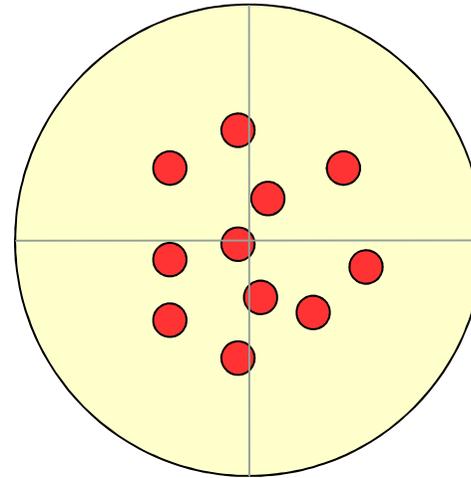


Biais et imprécision

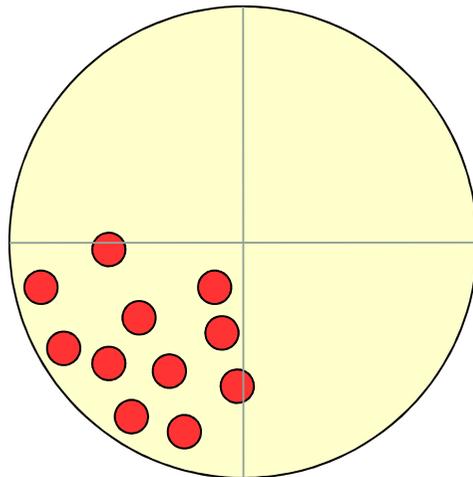
**Précis
mais biaisé**



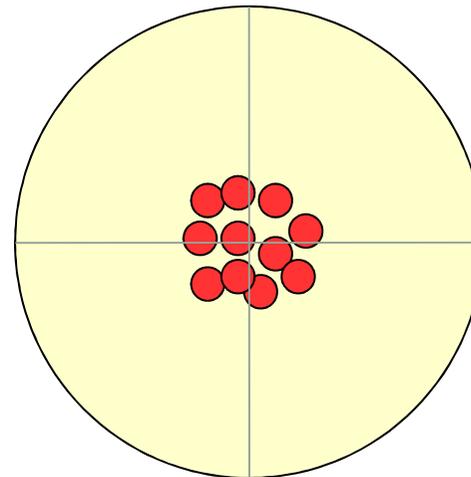
**Non biaisé
mais imprécis**



**Imprécis
et biaisé**



**Précis
et non biaisé**



Biais

- Généralités
 - Erreurs systématiques (contrairement aux fluctuations d'échantillonnage) → interprétation erronée
 - Principale limite des études épidémiologiques
→ essayer de les réduire / comprendre leur influence
- Les principaux :
 - Biais de sélection:
 - Échantillon non représentatif de la population
 - Biais de classement:
 - Exposition et/ou maladie mal définie(s)
 - Biais de confusion:
 - Relation biaisée par la présence d'un facteur influençant à la fois l'exposition et la maladie et non / mal pris en compte

Biais de sélection (1)

- Au moment de la constitution de l'échantillon: population d'étude non représentative de la population cible sur les paramètres étudiés
- Exemple 1: sur-représentation des malades ($\forall E$)

	E+	E-	Total
M+	100	100	200
M-	1400	3400	4800
Total	1500	3500	5000

→

	E+	E-	Total
M+	20	20	40
M-	140	340	480
Total	160	360	520

Population cible

Population source

$$P(M) = \frac{200}{5000} = 0,04$$

$$OR = \frac{ad}{bc} = 2,43$$

Sélection:
20% de M+
10% de M-

$$\tilde{P}(M) = \frac{40}{520} = 0,077$$

$$\tilde{OR} = \frac{ad}{bc} = 2,43$$

Biais de sélection (2)

- Au moment de la constitution de l'échantillon: population d'étude non représentative de la population cible sur les paramètres étudiés
- Exemple 2: sur-représentation des malades (**≠ selon E**)

	E+	E-	Total
M+	100	100	200
M-	1400	3400	4800
Total	1500	3500	5000

→

	E+	E-	Total
M+	20	20	40
M-	120	360	480
Total	140	380	520

Population cible

Population source

$$P(M) = \frac{200}{5000} = 0,04$$

$$OR = \frac{ad}{bc} = 2,43$$

Sélection:
20% de M+
10% de M-

**M- moins souvent
sélectionnés parmi les
E+ que parmi les E-**

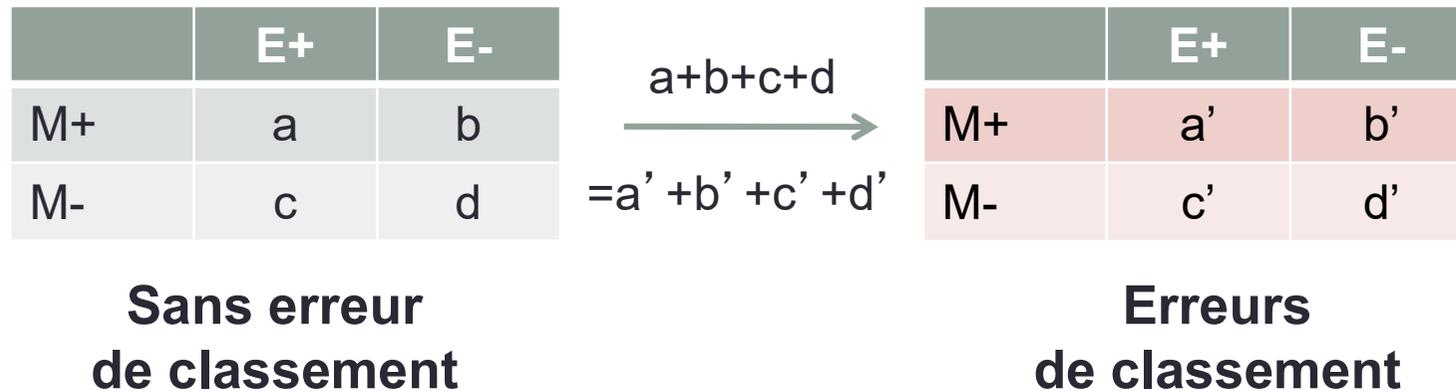
$$\tilde{P}(M) = \frac{40}{520} = 0,077$$

$$\tilde{OR} = \frac{ad}{bc} = 3,0$$

Biais de sélection

- C'est un problème certain en épidémiologie descriptive
- Peut être moins problématique dans les études analytiques, si le biais ne concerne pas à la fois l'exposition ET la maladie
- « Redressements » possibles: le principe est d'augmenter le poids des sujets sous-représentés dans l'échantillon

Biais de classement



- Erreurs non différentielles
 - Erreurs sur l'exposition indépendantes de la maladie (et réciproquement)
 - Sous-estimation systématique de OR
(pour une exposition en 2 classes)

- Erreurs différentielles:
 - Erreurs sur l'exposition affectent différemment les malades et les non malades (et réciproquement)
 - Modification de l'OR dans un sens ou dans l'autre

Exemple de biais de classement: le biais de mémoire

- Les malades se souviennent mieux de leurs expositions passées. Conséquence sur l'OR:

	E+	E-
M+	90	70
M-	1230	3460

**Sans erreur
de classement**

OR=3,62

$$\begin{array}{c}
 a+b+c+d \\
 \xrightarrow{\hspace{1cm}} \\
 =a' +b' +c' +d' \\
 = 4\ 850
 \end{array}$$

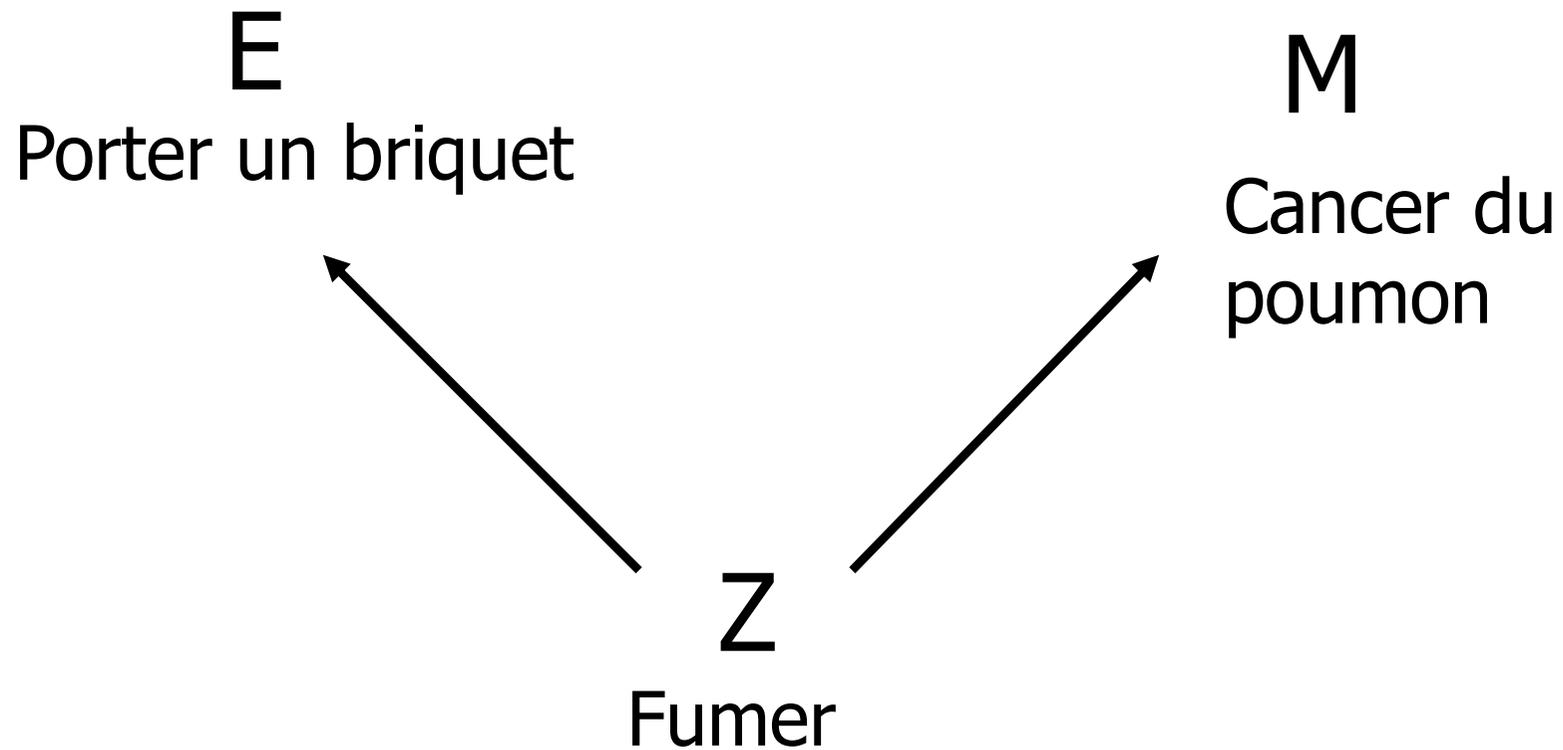
	E+	E-
M+	90	70
M-	984	3706

**Erreurs
de classement**

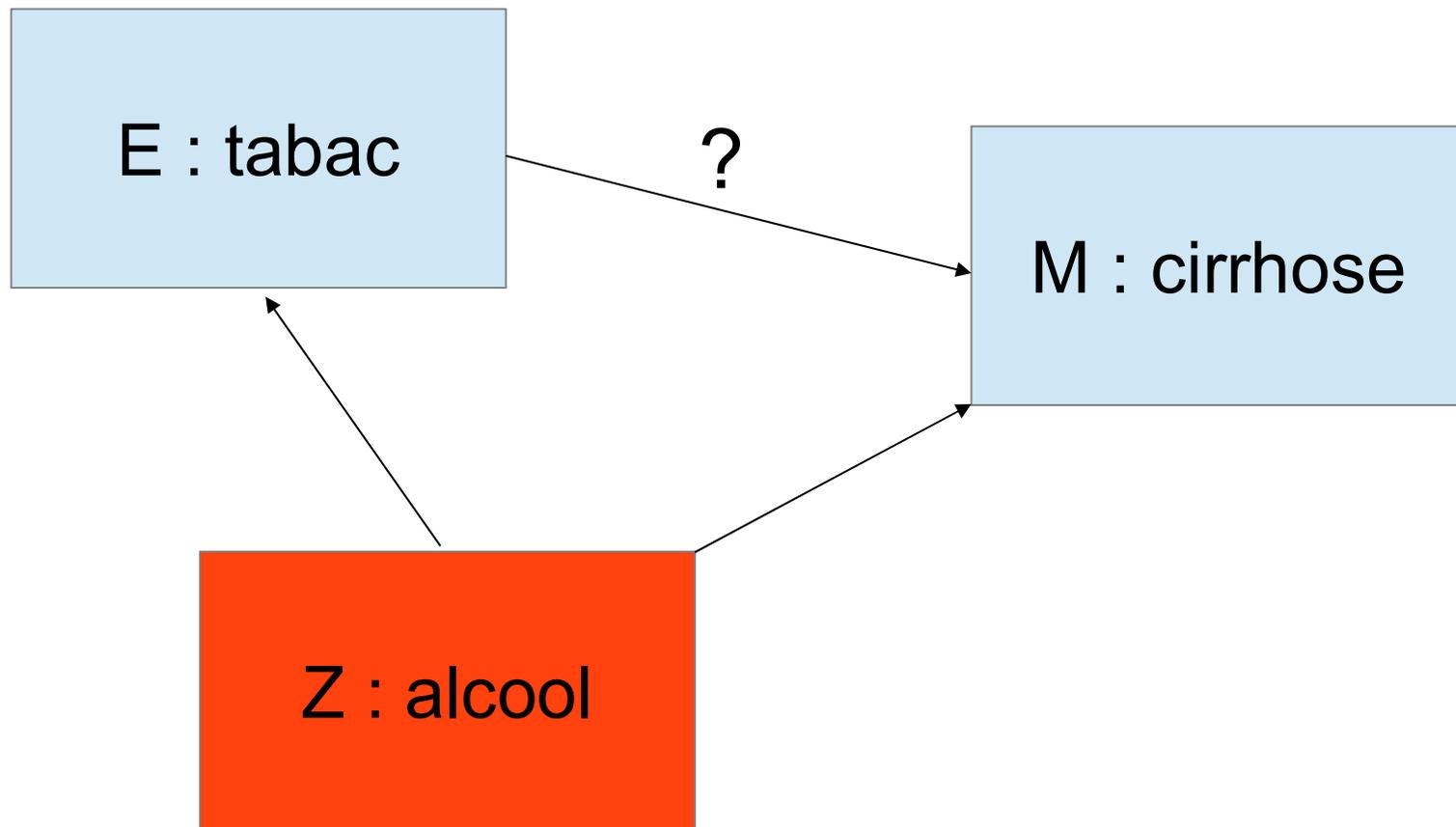
OR=4,84

Risque de sur-interprétation

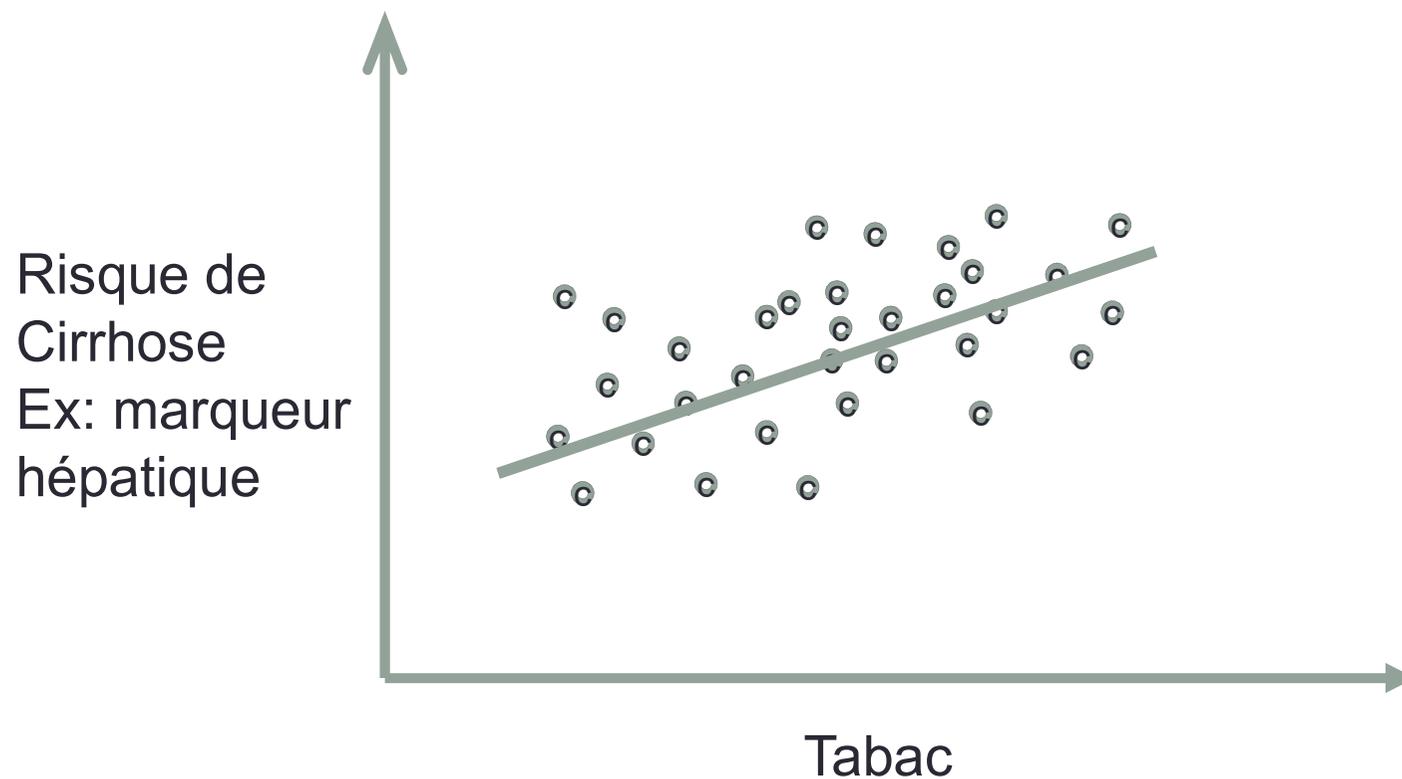
Biais de confusion: exemple



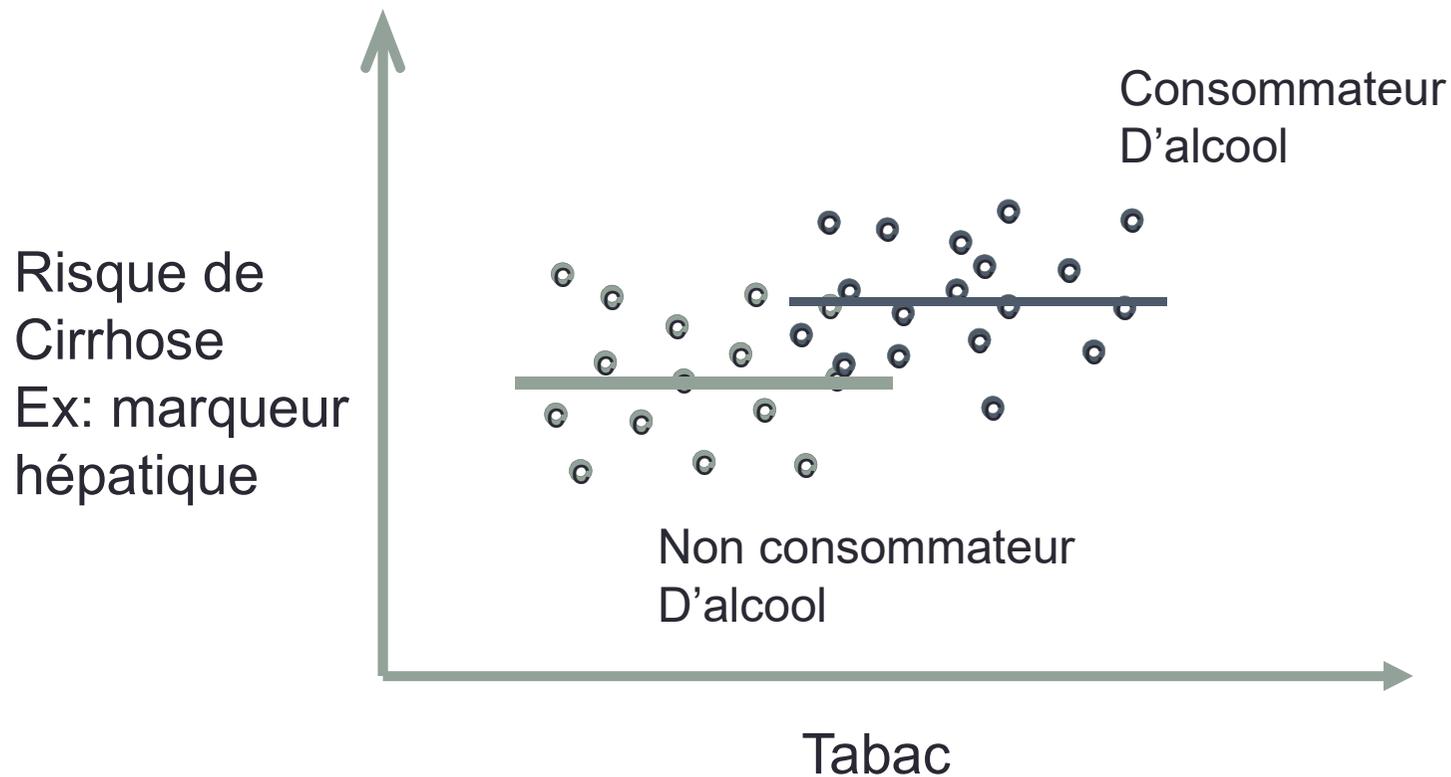
Biais de confusion schématiquement



Graphiquement (données fictives)



Graphiquement (données fictives)



Exemple chiffré

	M+	M-	Total
E+	200	600	800
E-	100	700	800
Total	300	1300	1600

M: accidents; E: chirurgiens

Incidences:

E+: 25%

E-: 12,5%

RR=2,0 [1,6 - 2,5]

Conclusion : risque 2 fois plus important avec le chirurgien

E+ qu'avec le chirurgien E-...

Question: E+ est-il dangereux??

Par classe de technique

Total	M+	M-	Total
E+	200	600	800
E-	100	700	800
Total	300	1300	1600

M: accidents; E: chirurgiens; Z: techniques

Z+
(davantage utilisée
dans les cas difficiles)

	M+	M-	Total
E+	170	230	400
E-	50	70	120
Total	220	300	520

Incidences:

E+: 42,5%

E-: 41,7%

RR=1,02 [0,8 – 1,3]

Z-

	M+	M-	Total
E+	30	370	400
E-	50	630	680
Total	80	1000	1080

Incidences:

E+: 7,5%

E-: 7,4%

RR=1,02 [0,7 – 1,6]

Conclusion : Indépendamment de la technique, les risques des deux chirurgiens ne sont pas significativement différents

La technique est facteur de confusion

Associations: (1) entre Z et E, (2) entre Z et M indépendamment de E

M: accidents; E: chirurgiens; Z: techniques

E+
(davantage utilisée
par E+ / E-)

	M+	M-	Total
Z+	170	230	400
Z-	30	370	400
Total	200	600	800

Incidences:

E+: 42,5%

E-: 7,5%

RR=5,6 [3,9 – 8,1]

E-

	M+	M-	Total
Z+	50	70	120
Z-	50	630	680
Total	100	700	800

Incidences:

E+: 41,7%

E-: 7,4%

RR=5,6 [4,0 – 7,9]

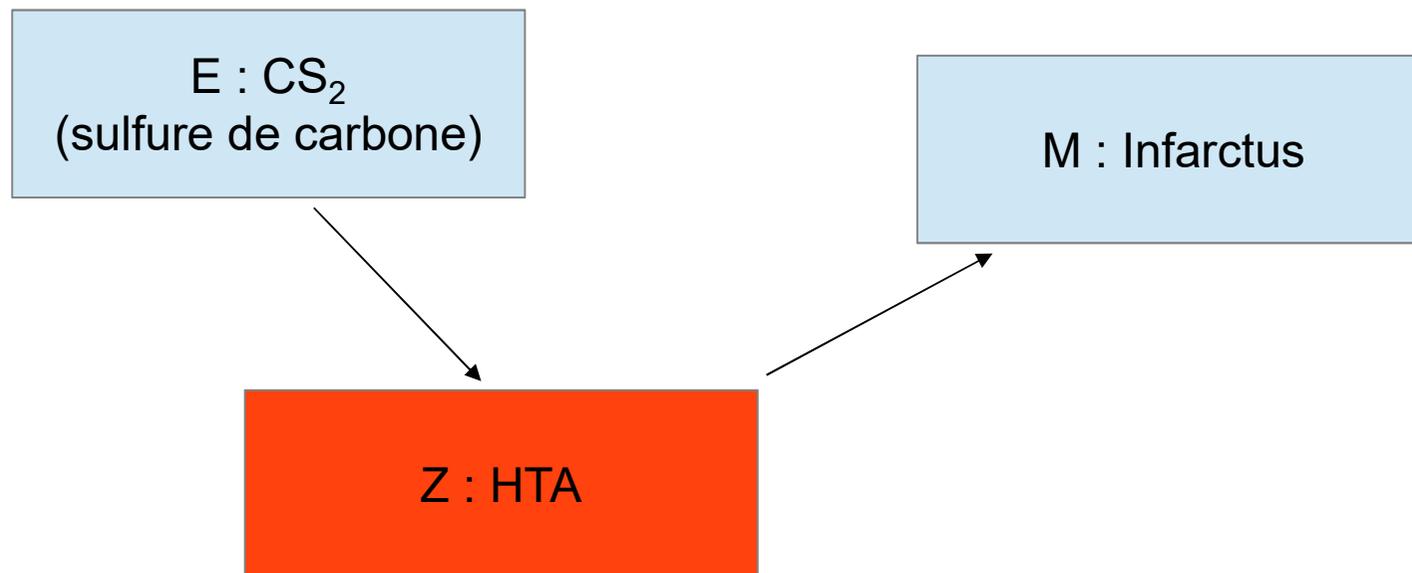
Biais de confusion

- Z est facteur de confusion si la relation brute entre E et M n'est pas la même que celle obtenue aux différents niveaux de Z
- Biais lié à la nature observationnelle des données
- Prise en compte:
 - Tirage au sort (essais cliniques)
 - Restriction de la population à une strate
 - Appariement: équilibrer la distribution de Z
 - Ex: jumeaux (Z=facteurs génétiques)
 - Cas-témoins (âge, sexe, CSP, etc.)
 - Modélisation

Exemple de modélisation

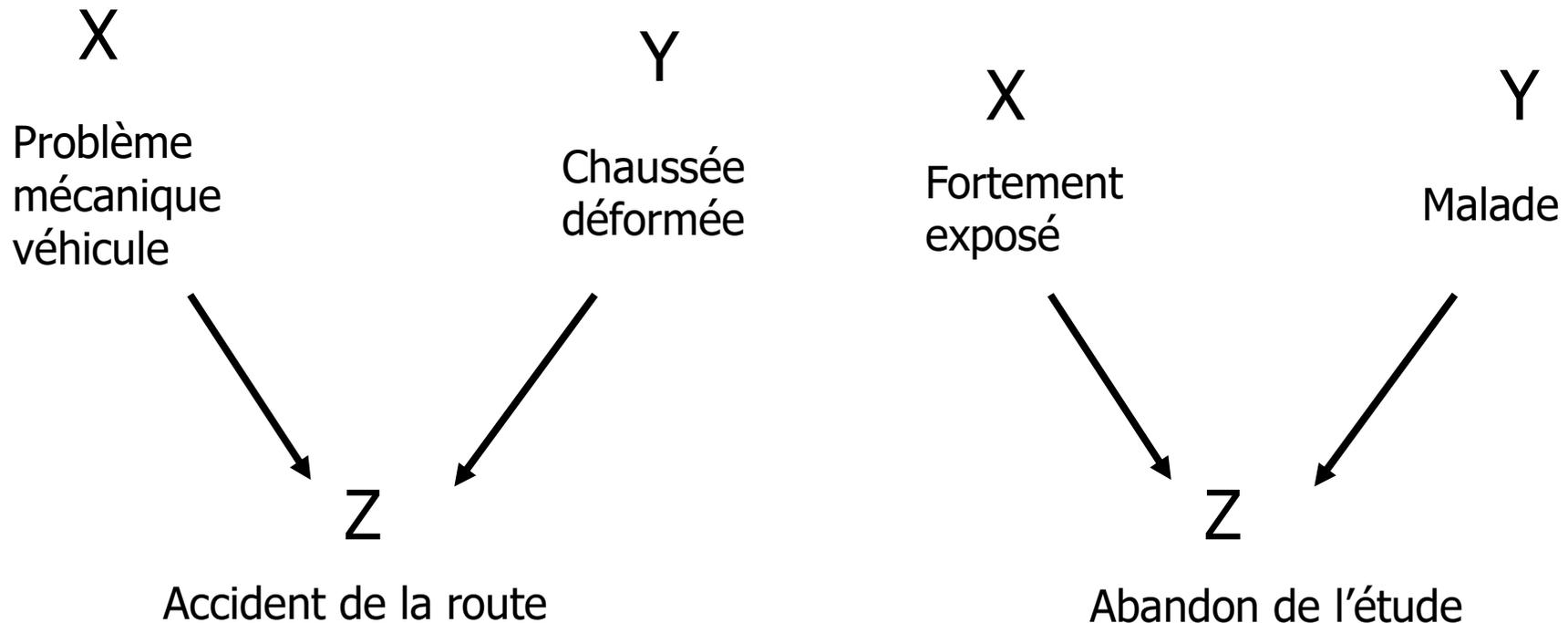
- Régression linéaire simple: $E(Y/X) = f(X) = a + b.X$
- Régression multiple: $E(Y/X,Z) = f(X,Z) = a + b1.X + b2.Z$
- Comparaison de Y selon X ajustée sur Z:
 - $E(Y/X=1,Z=0) - E(Y/X=0,Z=0) = (a+b1) - (a) = b1$
 - $E(Y/X=1,Z=1) - E(Y/X=0,Z=1) = (a+b1+b2) - (a+b2) = b1$
 - $b1$ représente donc l'effet de X quand Z est maintenu constant
 - Quel que soit Z (cela est imposé par le modèle)
 - Il est possible de complexifier le modèle pour permettre à $b1$ d'être différent selon Z

Remarque 1: facteur intermédiaire



- Il ne faut (généralement) pas ajuster sur un facteur intermédiaire

Remarque 2: les colliders (« collision »)



Il ne faut pas ajuster (ou restreindre) sur un collider

Remarque 3: interaction

Confusion :

pour chaque catégorie de Z, l'association entre E et M est similaire → le calcul d'un RR ajusté est possible

Interaction (= modificateur de l'effet) :

pour au moins une catégorie de Z, l'association entre E et M diffère → RR par strate de Z

exemple: métaboliseurs lents *versus* métaboliseurs rapides par CYP2D6

CAUSALITÉ

Notion fondamentale en Epidémiologie

Interprétations possibles des résultats

Test significatif

- On met en évidence une association statistiquement significative entre E et M:
 - Par ex, dans l'étude en question, les sujets les + exposés sont aussi ceux qui sont les + malades

- Se pose la question de la causalité
 - Présence d'un biais (ex: confusion)?
 - Fluctuation d'échantillonnage?

Nécessite généralement confirmation et des études complémentaires

- La relation est-elle cliniquement significative?

Test non significatif

- L'étude n'a pas mis en évidence de relation significative entre E et M:
- Faut-il pour autant écarter son existence ?
 - Manque de puissance: avec plus de sujets, peut-être l'aurait-on mis en évidence?
 - Fluctuation d'échantillonnage?
 - Biais?

La causalité

Difficile à prouver:

- Apport de l'expérimentation, mais pas toujours possible
 - considérations éthiques pour les études expérimentales chez l'homme
 - problème de l'extrapolation inter-espèce pour les études chez l'animal
- L'épidémiologie = observation de la population, donc:
 - Pas de conclusion causale directe
 - Mais des arguments (parfois forts) en faveur de la causalité
- La répétition d'études de qualité dans différentes populations et donnant des résultats concordants est importante

Éléments de discussion sur la causalité (critères de Bradford-Hill)

Internes à l'étude

- Séquence dans le temps (obligatoire)
- Force de l'association (niveau du RR)
- Spécificité de la cause et de l'effet (un seul facteur lié à la maladie)
- Relation type dose-effet: effet d'autant + grand que le niveau d'exposition augmente
- Cohérence interne à l'étude

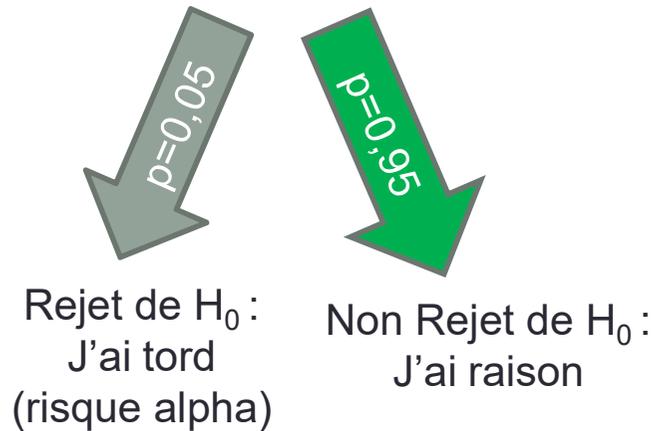
Externes à l'étude

- Constance de l'association
- Plausibilité biologique
- Cohérence avec les hypothèses de l'étude
- Parallélisme de la distribution dans l'espace et le temps

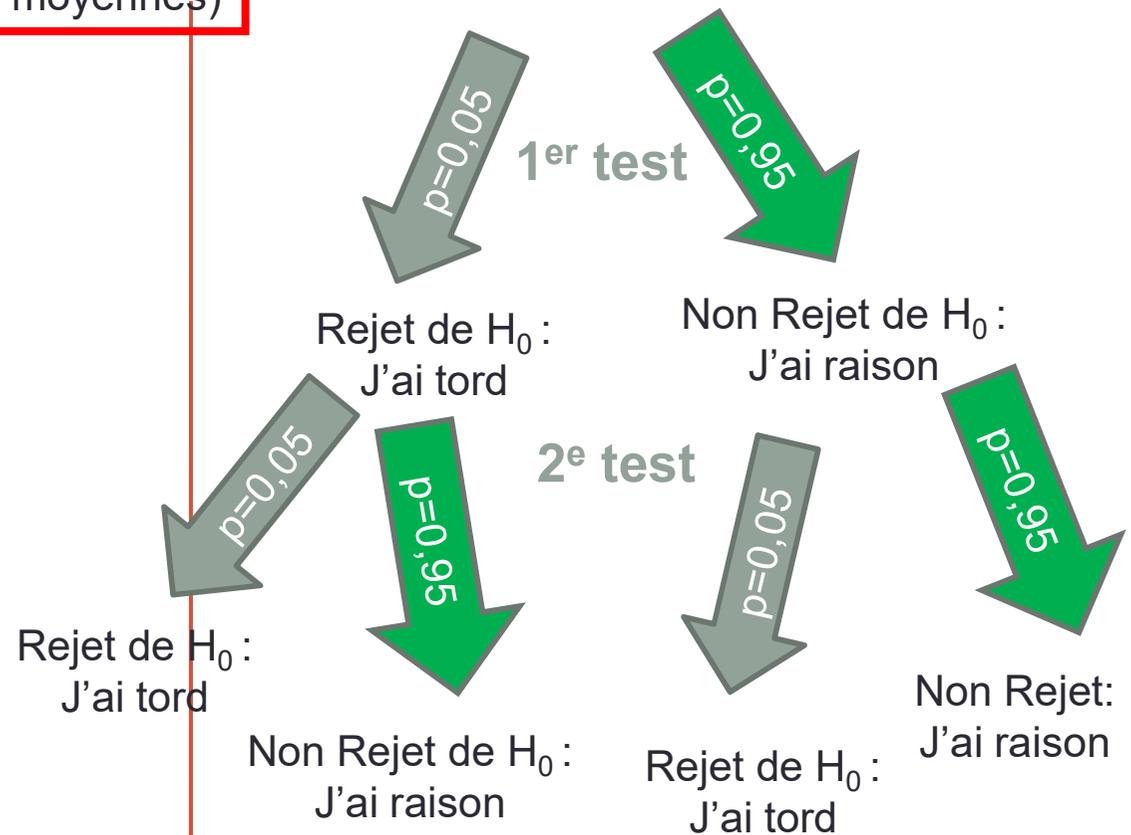
Attention à la multiplication des tests: on multiplie aussi le risque de se tromper!

Si hypothèse nulle H_0 vraie
(ex: pas de différence entre les moyennes)

1 test



2 tests



Remarque 1: pour n tests, $p(\text{tord})=1-0,95^n$
Remarque 2: par définition, sur 100 tests et en l'absence de différence, le hasard en rejette 5 (à tord!)

$$p(\text{raison})=0,95*0,95=0,9025$$

$$p(\text{tord})=1-p(\text{raison})=0,0975$$

Complémentarité avec la toxicologie

Toxicologie

- Expérimentation
- Animal/tissus
- Peu de sujets
- Mécanismes
- Causalité
- Fortes expositions possibles
- Exposition bien définie
- Avant l'introduction d'un produit

Epidémiologie

- Observation
- Homme
- Population
- Boîte noire
- Association
- Faibles niveaux d'exposition
- Exposition à évaluer
- Estimation du risque en population

Limites

Extrapolation inter-espèces
Extrapolation aux faibles doses

Biais
Evaluation de l'exposition
Extrapolation aux faibles doses

LECTURE D' UN ARTICLE

en guise d' exemple

Pharmacoépidémiologie

Pharmacoépidémiologie

- Définition: étude de l'utilisation et des effets (bénéfiques et indésirables) des médicaments sur de larges populations
- Limite des essais cliniques: sélection des patients, prescriptions selon procédure rigoureuse, relativement de courte durée
- Historique: malformations liées à la thalidomide, hypnotique utilisé chez les femmes enceintes (1960)

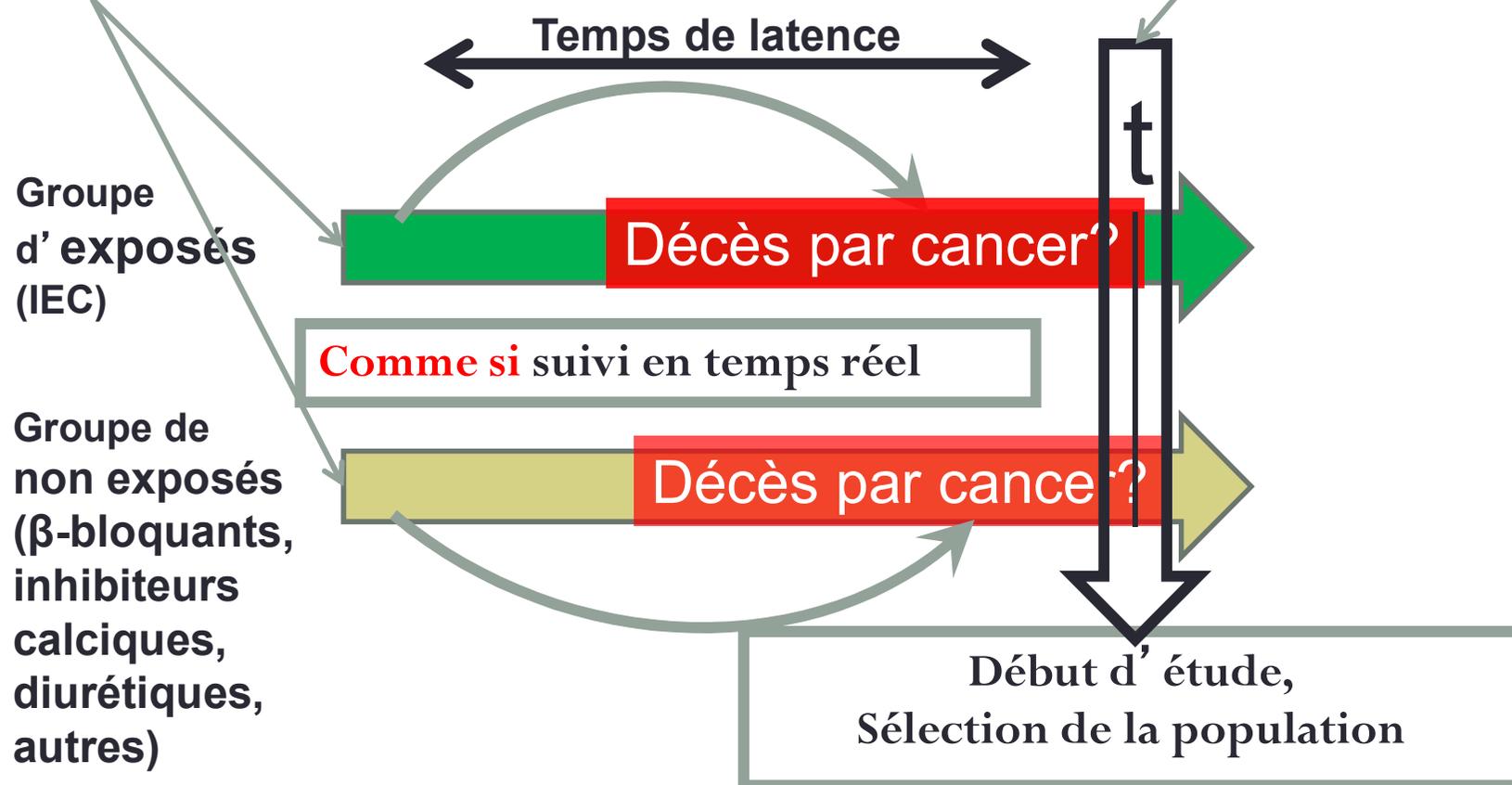
Ex: IEC et protection contre cancers?

- Objectif: étudier le risque de cancers chez les hypertendus traités par IEC (inhibiteur enzyme de conversion: classe thérapeutique d'antihypertenseurs)
- Méthodes: cohorte rétrospective de 5207 patients d'une clinique de Glasgow entre 1980 et 1995 – utilisation de registres des cancers
- Résultats : RR chez traités par IEC par rapport aux sujets traités par d'autres antihypertenseurs:
 - Cancer incident: $RR=0,67$ [0,50 – 0,90]
 - Mortalité par cancer: $RR=0,63$ [0,42 – 0,96]
- Conclusion: l'utilisation d'IEC à long terme pourrait protéger du risque de cancers; ce n'est qu'une hypothèse à vérifier par des études expérimentales

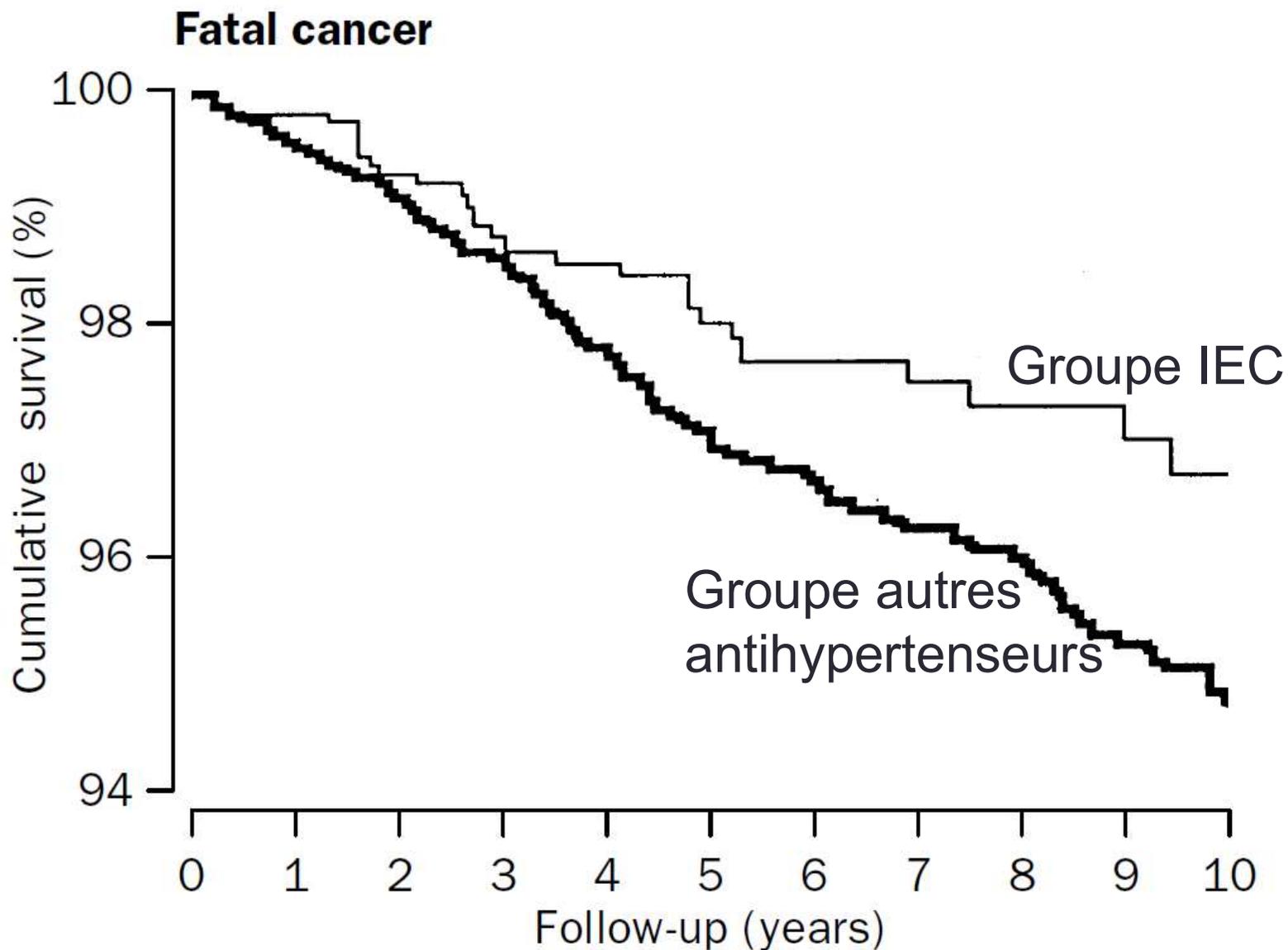
Etude=cohorte rétrospective

Première prescription d'un antihypertenseur ou 1^{er} janvier 1980

Décès ou 31 décembre 1995



Résultats



Quelques remarques

- Validation des cas de cancers par l'utilisation de registres: faites « à l'aveugle » de la connaissance du traitement (éviter un biais de classement)
- De même, classement des traitements « à l'aveugle » de la connaissance sur l'apparition ultérieure de cancer
- Prise en compte de facteurs de confusion: l'âge, le sexe, l'année d'observation et le statut tabagique
- Même risque relatif chez les fumeurs et les non-fumeurs (pas d'interaction)
 - Risque calculé sur la population entière est le même que celui calculé séparément chez les fumeurs et non-fumeurs, le tabac n'est donc pas un facteur de confusion

Causalité?

- En faveur:
 - Il existe une plausibilité biologique: par leur effet antihypertenseur, diminution de composés intervenant dans la croissance cellulaire nécessaire au développement de cancers
 - Relation non pas liée à la diminution de la pression artérielle, ni au tabac
 - Discussion sur les différents types de biais
- En défaveur:
 - Nombreux tests réalisés avant de trouver une association significative

Conclusion

- Que faire à l'issue de cette étude?
 - 1) Prescription d' IEC pour prévention des cancers?
NON!
 - 2) Etude(s) de confirmation?
Oui!
- L' étude a servi à générer une hypothèse qu'il faut vérifier
 - Risque de fluctuation d'échantillonnage demeure
 - Tests multiples
- Ici, l'effet protecteur n'a pas été confirmé

FIN – ACTUALITÉ

En résumé (ce qu'il faut retenir)

- Savoir définir la discipline (avec vos propres mots)
- Savoir reconnaître un type d'enquête
 - Transversal
 - Cas-témoins
 - Cohorte
- Notions statistiques:
 - Estimations de paramètres au niveau de populations;
 - Fluctuations d'échantillonnage et biais;
 - Tests=prise en compte de la variabilité pour prise de décision.
- Indicateurs de risque: descriptifs et étiologiques (RR et OR)
- Connaître les 3 principaux types de biais (sélection, classement, confusion) et le contexte dans lequel ils peuvent apparaître
- Interprétation des résultats statistiques dans leur contexte

COVID-19

- Propagation du virus
 - Comptage
 - Modélisations → détermination du R, déterminants
- Epidémiologie descriptive
 - Age des malades / décès
 - Catégories de la population
 - Variabilité des symptômes
 - Evolution des consommations de soin
- Epidémiologie analytique
 - Hydroxychloroquine: épidémiologie, recherche clinique, méta-analyses
 - Anti-hypertenseurs
 - Facteurs associés au risque de forme grave de COVID-19
 - Efficacité et sécurité des vaccins