

# Epidémiologie

## Plan

---

1. Introduction
2. Enquête de cohorte
3. Enquête cas-témoins
4. Mesures de risques
5. Mesures d'association
6. Biais de sélection
7. Biais de classement
8. Biais de confusion
9. Stratégie d'analyse
10. Puissance
11. Modèles multivariés
12. Régression logistique

**Philippe SAINT PIERRE**

Université Paul Sabatier – Toulouse III

Institut de Mathématiques de Toulouse

[philippe.saint-pierre@math.univ-toulouse.fr](mailto:philippe.saint-pierre@math.univ-toulouse.fr)

# Epidémiologie

## 1. Introduction

**Philippe SAINT PIERRE**

Université Paul Sabatier – Toulouse III

Institut de Mathématiques de Toulouse

[philippe.saint-pierre@math.univ-toulouse.fr](mailto:philippe.saint-pierre@math.univ-toulouse.fr)

# 1. Introduction

---

I. Epidémiologie

II. Différents types d'enquêtes

III. Population cible et population source

IV. Fluctuations d'échantillonnage

V. Biais en épidémiologie

# I. Epidémiologie

---

- **Historique**

Étude des épidémies (maladies transmissibles, comptage)



Recherche des causes (variations géographiques, temporelles, entre groupes)



Généralisation aux maladies non contagieuses



Étude de la distribution des maladies et des facteurs qui influencent cette distribution

Étude de l'étiologie des pathologies

- **Recherche observationnelle analytique (ou étiologique)**

- Rechercher l'**existence d'une relation** entre une exposition et la survenue d'une maladie (**mesure de risque**) et mesurer cette relation (**mesure d'association**)

- Un facteur E est un **facteur de risque** pour une maladie M si l'exposition à E modifie la vitesse d'apparition de la maladie M

- Notion et critères de **causalité**

# Quelques questions en épidémiologie

---

- **Question I** : Quelle est la "fréquence" d'une pathologie ou d'un comportement?
  - **Ex** : cancers du sein, pratique des mammographies (selon l'âge, le sexe ou la région géographique)
    - ↳ Décrire et quantifier un phénomène de santé
- **Question II** : Comment évolue la fréquence d'une pathologie ou d'un comportement au cours du temps ?
  - **Ex** : usage du préservatif entre 1980 et 2000, épidémie de sida en France entre 1980 et 2000.
    - ↳ Surveiller les tendances évolutives, mettre en place des systèmes d'alerte
- **Question III** : Comment varie la fréquence d'une pathologie d'un endroit à un autre?
  - **Ex** : variation géographique de la fréquence de l'infarctus du myocarde
    - ↳ Corrélations écologiques

EPIDEMIOLOGIE DESCRIPTIVE

⇒ générer des hypothèses

# Quelques questions en épidémiologie

---

- Question IV : Peut-on identifier des facteurs associés à une maladie ou à un comportement donné ?
  - Ex : facteurs associés au cancer du poumon

↳ Facteurs d'exposition  $\iff$  Phénomène étudié (Recherche analytique)

- Question V : Y a-t-il une relation causale entre un facteur d'exposition et une maladie ou un comportement donné ?
  - Ex : l'infection à HPV est-elle susceptible de provoquer la survenue de cancer du col de l'utérus ?

↳ Facteurs d'exposition  $\implies$  Phénomène étudié (Recherche analytique)

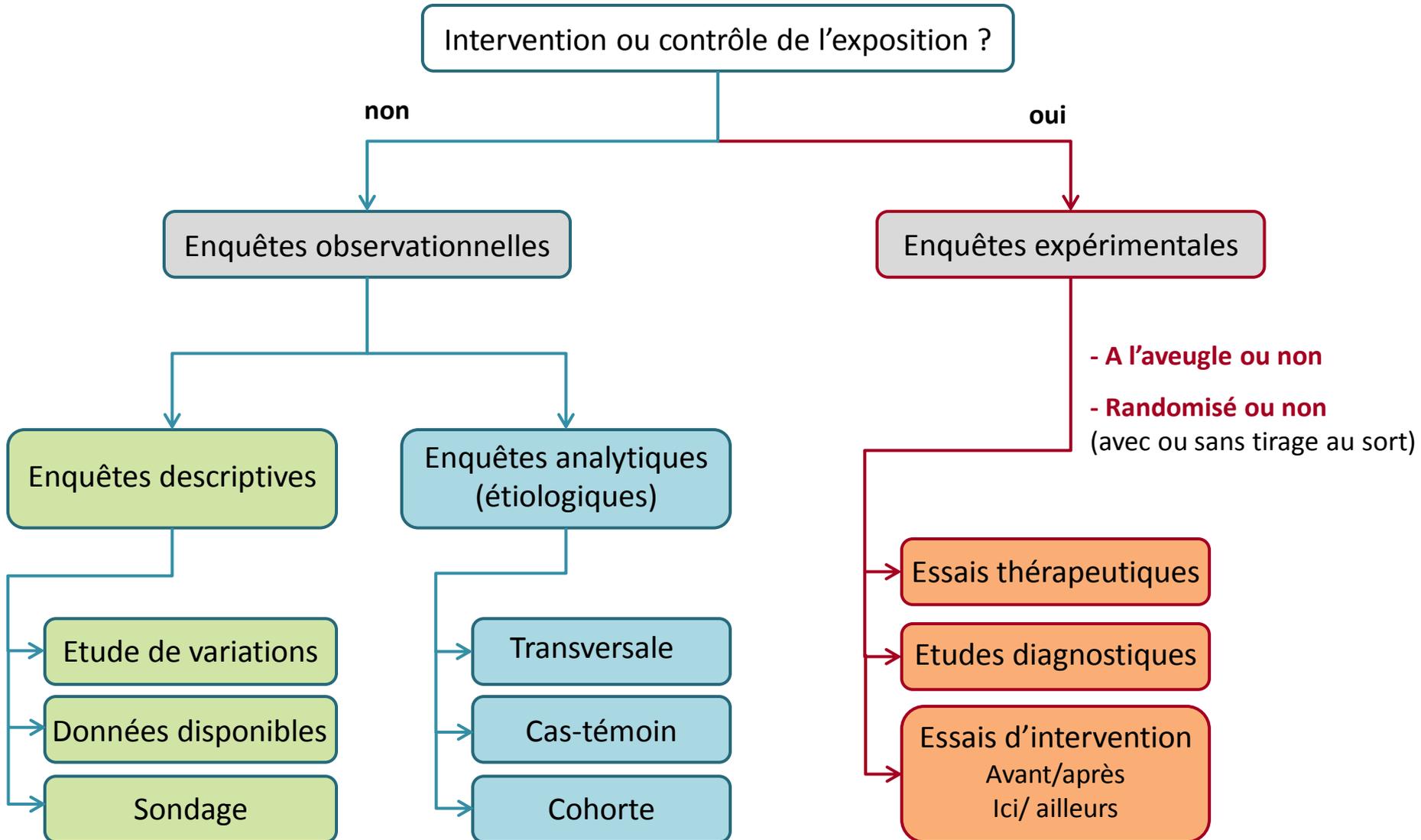
EPIDEMIOLOGIE ANALYTIQUE

- Question VI : Quelle est l'efficacité d'une intervention sur un facteur d'exposition ou d'un traitement pour diminuer la survenue d'une maladie ou modifier un comportement ?
  - Ex : Quelle est l'efficacité d'un programme de dépistage du cancer du sein dans une population sur la mortalité par ce type de cancer ?

↳ Recherche expérimentale, évaluation d'intervention

EPIDEMIOLOGIE EVALUATIVE

# II. Différents types d'enquêtes



# Enquête descriptive

---

- Étude de la fréquence et de la **répartition** des **paramètres de santé** ou des **facteurs de risque** dans une population.
- **Études de variations**
  - **Variations dans le temps**
    - Ex : Augmentation de la fréquence de certaines maladies permet de soupçonner l'introduction progressive d'un nouvel agent toxique (Tabac, amiante,...).
  - **Variations géographiques**
    - Ex : Moins de cancer du sein au Japon qu'en Europe (facteur génétique ou environnementaux ?) => Étude de migration (taux identiques entre femmes américaines et descendances d'immigrants Japonais) => facteurs environnementaux
  - **Les « clusters » (agrégats spatio-temporels)**
    - Plusieurs cas d'une maladie rare en des lieux proches et dans un intervalle de temps court (agents cancérigènes, ...)

# Enquête descriptive

---

- **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, ...)
- **Enquêtes peu coûteuses et rapides** permettant d'évaluer une hypothèse ou d'alimenter un faisceau d'arguments

# Enquête expérimentale

---

- Le chercheur a le **contrôle complet de l'exposition** (la nature précise de l'exposition, le moment de son utilisation, du choix des sujets qui vont en bénéficier et du déroulement de l'essai).

↳ Objectif : éviter l'apparition de différences autres que l'effet de l'exposition.

- Le statut de l'exposition peut être

- Connu du patient
- **A l'aveugle** : le patient ne connaît pas son exposition
- **En double aveugle** : le patient et l'examineur ne connaissent pas l'exposition

↳ Objectif : éviter les **biais d'interprétation** (fonction de la conviction des protagonistes)

- L'expérience peut être randomisée ou non randomisée

- **Non randomisé** : les groupes traités et témoins sont choisis par l'investigateur.
- **Randomisé** : les groupes comparés sont constitués par tirage au sort

↳ Objectif : éviter que l'attribution du traitement soit influencée par l'état de santé (**biais de sélection**) ou par une autre caractéristique (**biais confusion**).

# Enquête expérimentale

---

- **Ex : essai thérapeutique** permet d'évaluer l'effet d'un traitement, l'exposition est le placebo ou le traitement
- **L'essai randomisé à l'aveugle** est la meilleure méthode d'évaluation
  - ➔ **permet de montrer la causalité !**  
(si ce n'est pas le cas, le niveau de preuve est plus faible)
- **Limites**
  - La **randomisation n'est pas toujours possible** pour des raisons éthiques  
Ex : effet du tabac, évaluation d'une prévention, ...
  - Les **conditions de l'expérience** ne sont **pas toujours vérifiées** dans la pratique

# Enquête quasi-expérimentale

---

- **Enquête "avant-après"**
  - comparaison les situations avant et après une intervention
  - Les sujets peuvent être leur propre témoin
  - **Ex** : étude sur le comportement vis à vis du tabagisme avant et après une campagne de prévention contre le tabac.
  - **Difficulté** : variation spontanée des indicateurs qui se serait produite même en l'absence de l'intervention (mise en place au même moment d'autres mesures de santé, changements socioculturels, ...)
- **Enquête "ici-ailleurs"**
  - Comparaison, au même moment, des groupes distincts géographiquement où l'un reçoit l'intervention et l'autre pas
  - **Ex**: deux services hospitaliers pratiquant ou non l'intervention
  - **Difficulté** : possibilité d'une différence initiale entre les deux populations comparées

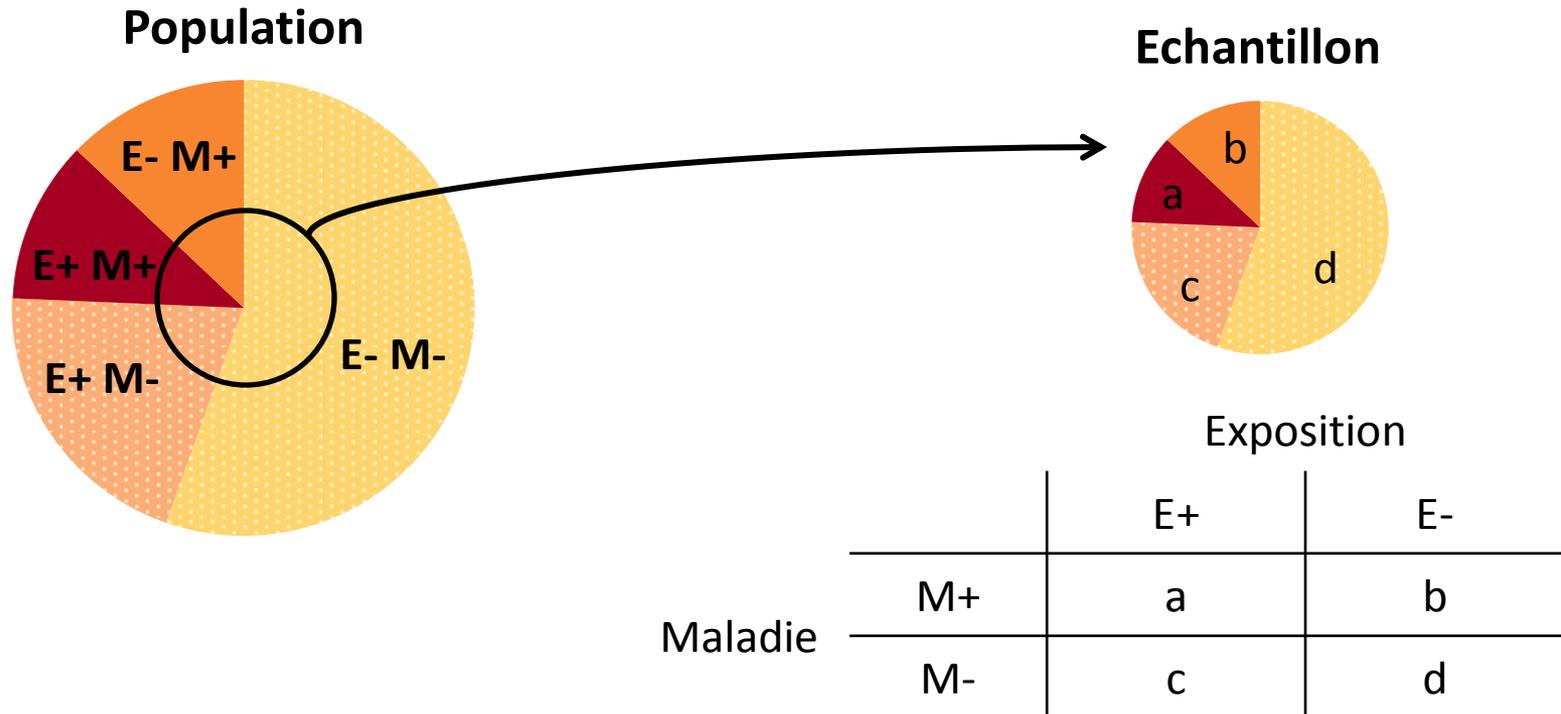
# Enquête analytique

---

- Enquête sur un **échantillon de sujets** veillant à recueillir des **informations individuelles** sur la maladie et les expositions
  - Sélection d'un échantillon de la population cible et mesures d'indicateurs adaptés
  - Etude des associations à un niveau individuel
  - Prise en compte de facteurs de confusion
  - Prise en compte de la chronologie des évènements
- 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 analytique : transversale

- Etude d'un **échantillon représentatif** de la population étudiée : pas de sélection sur l'exposition (cohorte) ou la maladie (cas-témoins)
- **Objectif** : étudier l'association entre l'exposition et la maladie
- **Image instantanée d'un phénomène**: recueil des informations au moment de l'inclusion



# Enquête analytique : transversale

---

- **Avantages**

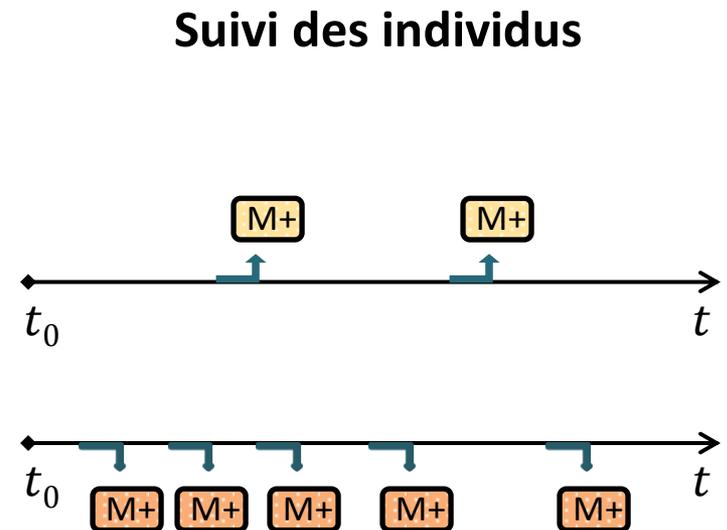
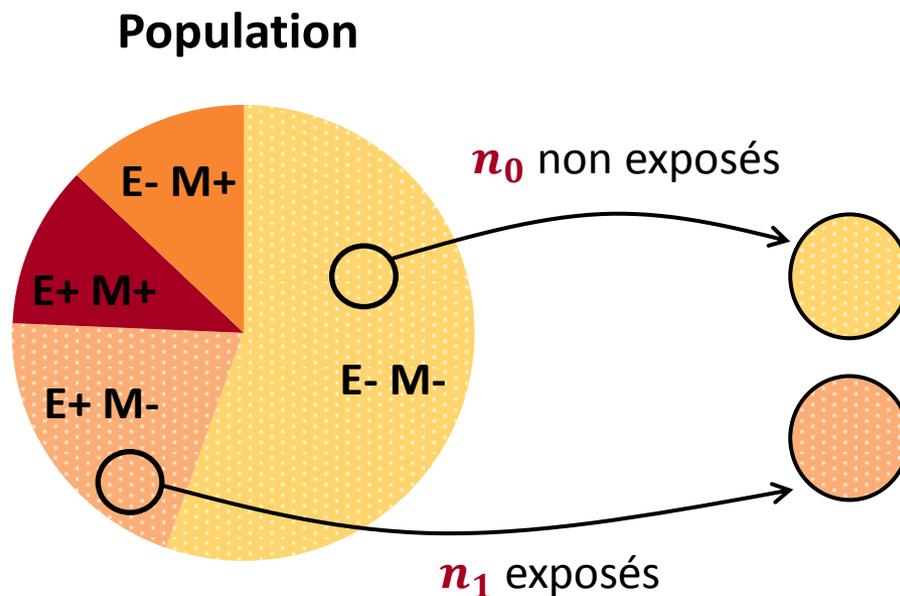
- Enquête facile à mettre en œuvre, rapide et peu coûteuse

- **Limites**

- Pas adapté aux maladies rares et aux expositions rares
- Biais de sélection possible dans la sélection de l'échantillon
- Chronologie des événements E-M (séquence temporelle) pas évidente
- Recueil de l'exposition souvent rétrospectif (biais de mémorisation)
- Enquête observationnelle avec le plus faible niveau de preuve

# Enquête analytique : cohorte

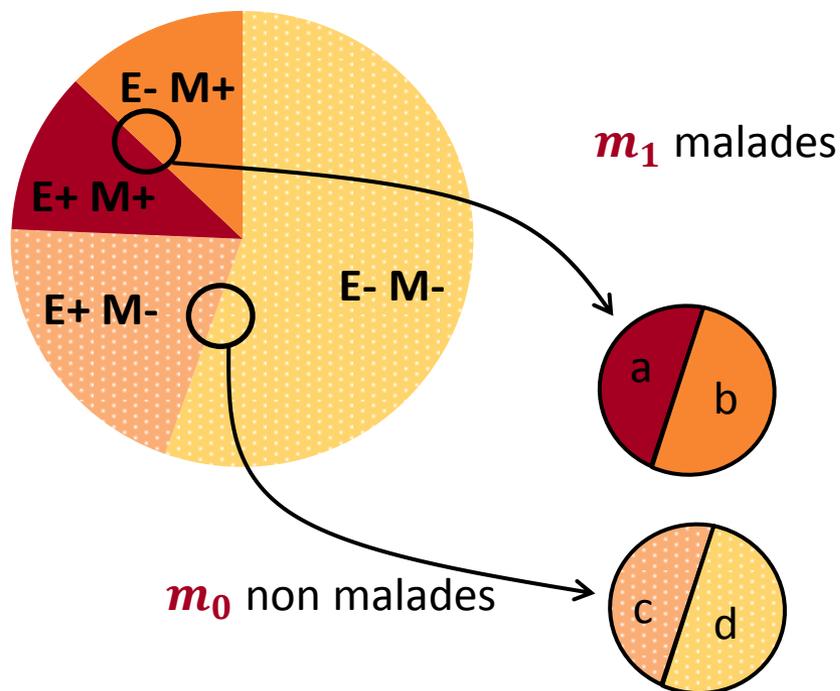
- **Cohorte** : suivi dans le temps (étude longitudinale), selon un protocole pré-établi, d'un groupe de sujets qui répondent à une définition donnée
- Sélection d'un **échantillon de non malades** :  $n_1$  exposés et de  $n_0$  non-exposés
- **Objectif** : comparer la survenue de la maladie chez les E+ et les E-



# Enquête analytique : cas-témoins

- Sélection d'un échantillon de  $m_1$  malades (cas) et de  $m_0$  non-malades (témoins)
- Objectif : comparer la répartition de facteurs de risque entre les cas et les témoins
- Recueil de l'exposition est toujours rétrospectif

## Population

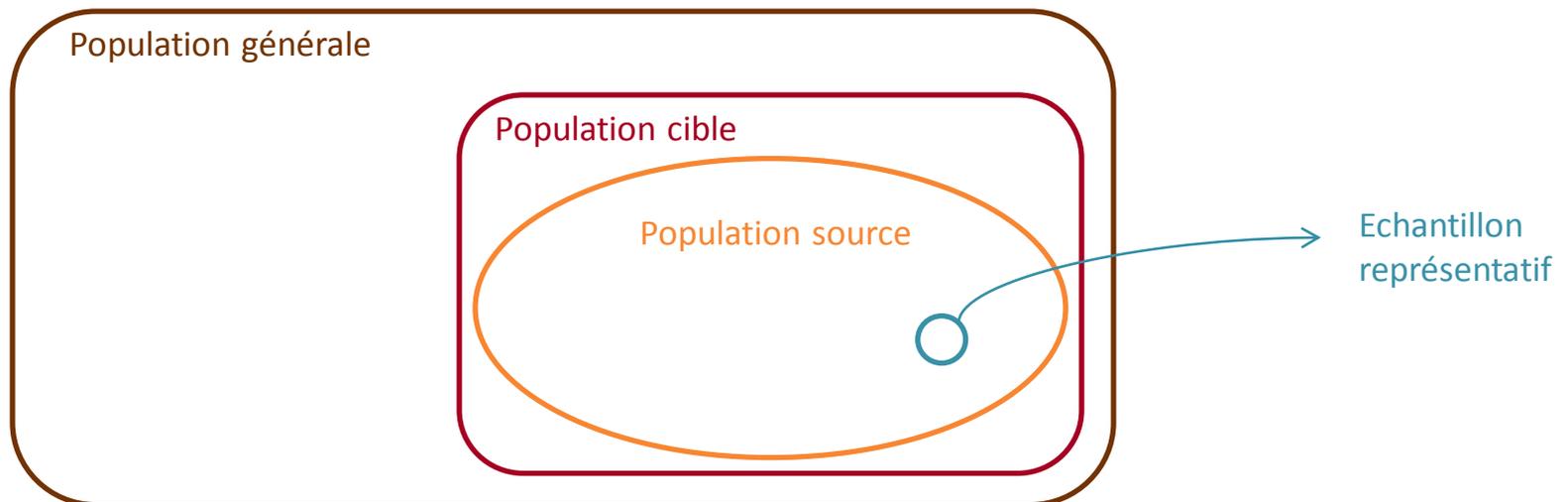


|         |    | Exposition |    |       |
|---------|----|------------|----|-------|
|         |    | E+         | E- |       |
| Maladie | M+ | a          | b  | $m_1$ |
|         | M- | c          | d  | $m_0$ |

# III. Population cible et population source

---

- La **population cible** est l'ensemble des sujets visés par l'enquête, c'est-à-dire la population pour laquelle **les résultats de l'enquête pourront être généralisés**
- La **population source** est l'ensemble des sujets à partir desquels on va constituer l'échantillon. La population source doit être **voisine de la population cible** et généralement plus facile d'accès parce que rassemblée (entreprise, hôpital, école)
- La **population de l'enquête** est soit la population source (exhaustivité) soit un **échantillon représentatif** de la population source



# IV. Fluctuations d'échantillonnage

---

- Les enquêtes épidémiologiques sont souvent réalisées à partir d'un échantillon de la population source
  - L'échantillon doit être représentatif de la population source
  - L'échantillonnage par tirage au sort est la seule méthode qui permet d'obtenir un échantillon représentatif à partir d'une liste exhaustive de la population source

Difficultés: - aucune liste n'est exhaustive en population générale  
- risque de biais de sélection liés aux non-réponses

- La sélection d'un échantillon représentatif est sujette a une erreur liée aux fluctuations d'échantillonnage

➔ erreur aléatoire (imprévisible)

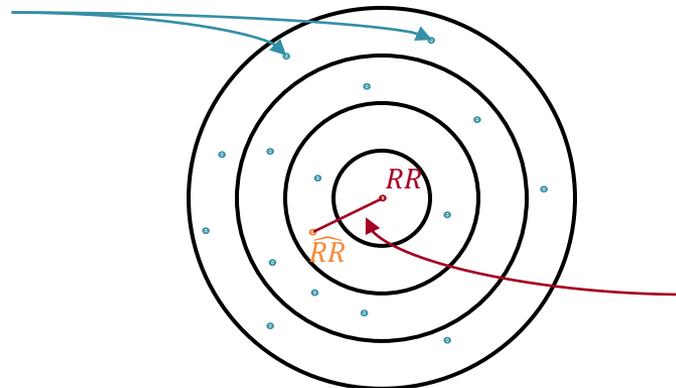
# Fluctuations d'échantillonnage

- Les mesures de risques (prévalence, incidence, ...) et d'associations (RR, OR...) sont des **estimations** obtenues à partir de ces échantillons
- Soit  $\widehat{RR}$  estimateur sans biais de  $RR$   $\longrightarrow E(\widehat{RR}) = RR$   
 $\longrightarrow$  les écarts entre  $\widehat{RR}$  et  $RR$  sont nuls en moyenne

Observations

$RR$  (vraie valeur)

$\widehat{RR}$  (estimation)



Fluctuation d'échantillonnage :  
"Imprécision autour de  $RR$ "

- **Erreur aléatoire** entre  $\widehat{RR}$  et  $RR$  prise en compte dans **les tests statistiques**
- Les estimations doivent être données avec **leur intervalle de confiance à 95%**.  
 $\longrightarrow$  intervalle dans lequel la vraie valeur a de fortes chances (95%) de se situer
- **Erreur diminue** quand la taille de **l'échantillon augmente**

# V. Biais en épidémiologie

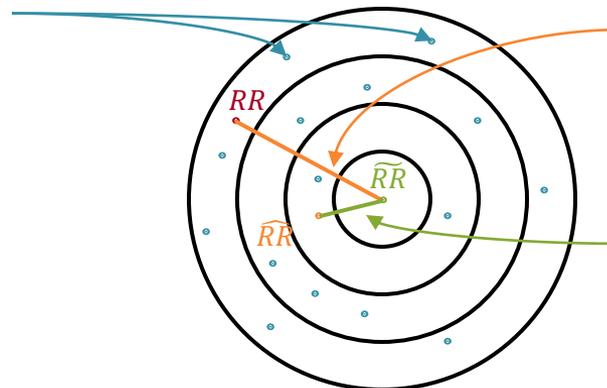
- Un biais est une **erreur systématique** qui fausse les résultats dans un sens donné.
- Erreur qui **se répète sur chaque échantillon**
- Le biais **ne diminue pas** avec la taille de l'échantillon
- Soit  $\widehat{RR}$  estimateur biaisé de  $RR$   $\longrightarrow E(\widehat{RR}) = \widetilde{RR} \neq RR$   
 $\longrightarrow$  les écarts entre  $\widehat{RR}$  et  $RR$  ne sont pas nuls en moyenne

Observations

$RR$  (vraie valeur)

$\widehat{RR}$  (estimation)

$\widetilde{RR} = E(\widehat{RR})$



Biais : "Erreur de visée"

Fluctuation d'échantillonnage :  
"Imprécision autour de  $\widetilde{RR}$ "

# Biais en épidémiologie : 3 types de biais

---

- **Biais de sélection** : erreur liée à la **sélection des sujets** à inclure et au suivi des sujets : problèmes de représentativité ou de comparabilité des groupes
  - **Echantillon non représentatif**
  - **Vision déformée de la population source**
- **Biais de classement** : erreur due à une **mesure erronée de l'exposition ou de la maladie** (subjectivité de l'enquêteur, mauvais outil de mesure, biais de mémorisation, ...)
  - **Echantillon représentatif mais classement erroné**
  - **Vision déformée de la réalité**
- **Biais de confusion** : erreur due à la présence d'un tiers facteur lié à l'exposition et à la maladie
  - **Echantillon représentatif mais difficultés à observer (isoler) la relation entre E et M** (même problème dans la population cible)
  - **Ce n'est pas un biais au sens statistique**
  - **Ex : le tabagisme peut être un facteur de confusion dans la relation entre la consommation d'alcool et le risque de cancer du poumon.**

# Biais en épidémiologie

---

- **Biais différentiel**
  - Les erreurs affectent **différemment** les groupes comparés.
  - Peut entraîner une surestimation ou une sous-estimation de l'association (RR, OR, ...)
  - Ex : diagnostic non objectif de la maladie en connaissant le statut sur l'exposition
- **Biais non différentiel**
  - Les erreurs affectent **indifféremment** les groupes comparés
  - Entraîne toujours une **diminution de l'association** (RR se rapproche de 1)
  - Ex : erreur systématique sur un appareil de mesure nécessaire au diagnostic
- **Description d'un biais** : le type, différentiel ou non, la direction et son importance
- **Contrôle des biais au moment de la planification ou de l'analyse statistique**

# Plan

---

2. Enquête de cohorte
3. Enquête cas-témoins
4. Mesures de risques
5. Mesures d'association
6. Biais de sélection
7. Biais de classement
8. Biais de confusion
9. Stratégie d'analyse
10. Puissance
11. Modèles multivariés
12. Régression logistique

# Epidémiologie

## 2. Enquête de cohorte

**Philippe SAINT PIERRE**

Université Paul Sabatier – Toulouse III

Institut de Mathématiques de Toulouse

[philippe.saint-pierre@math.univ-toulouse.fr](mailto:philippe.saint-pierre@math.univ-toulouse.fr)

# 2. Enquête de cohorte

---

## I. Définitions d'une cohorte

## II. Cohorte étiologique : enquête exposé – non exposé

## III. Méthodologie générale

- Recueil de données, mode de recrutement des sujets, choix du groupe de référence, prévalence de l'exposition

## IV. Mise en place d'une enquête exposé-non exposé

- Constitution de la cohorte, recueil des informations
- Suivi, perdus de vue et biais dans une enquête de cohorte

## V. Avantages et limites

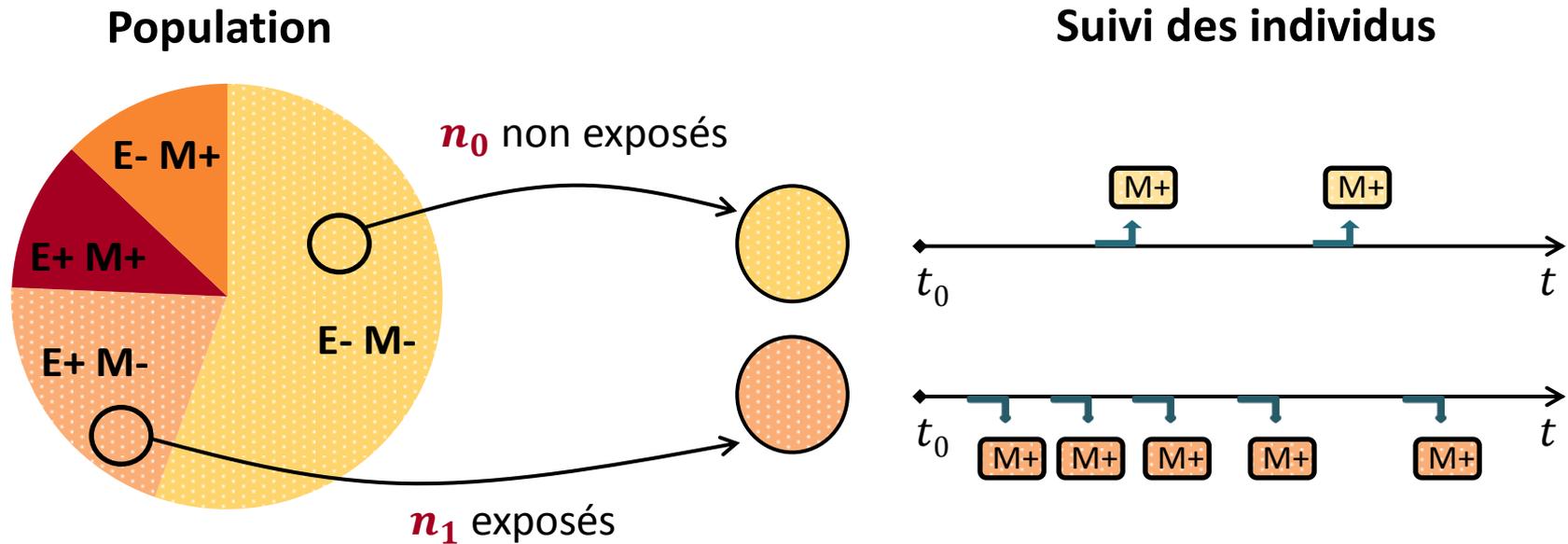
# I. Définitions d'une cohorte

---

- Définitions : (1) division de la légion romaine, (2) groupe
- Définition (Epidémiologie) : **Suivi dans le temps (étude longitudinale)**, selon un protocole pré-établi, d'un groupe de sujets qui répondent à une définition donnée :
  - exposé-non exposé : cohorte étiologique
  - sélectionnés en population générale : cohorte descriptive
  - ou souffrant d'une pathologie particulière
- **L'objectif des enquêtes de cohorte** est de mesurer et de comparer la survenue d'une pathologie dans une population en fonction de l'exposition à un ou plusieurs facteurs prédictifs.

## II. Cohorte étiologique : exposé – non exposé

- Sélection d'un **échantillon de non malades** :  $n_1$  exposés et de  $n_0$  non-exposés



- Ou sélection d'un **échantillon représentatif** et suivi des individus non malades

# Cohorte étiologique : exposé – non exposé

---

- **Objectif** : épidémiologie observationnelle analytique
  - ➡ Rechercher l'existence d'une relation entre une exposition et la survenue d'une maladie et mesurer l'intensité de cette relation
- **Déroulement** se rapproche des études expérimentale
  - Recrutement des sujets **non malades** exposés ou non exposés au facteur étudié
  - Les sujets sont **suivis dans le temps** et chaque nouveau cas de la maladie est enregistré
  - Estimation et comparaison des **incidences de la maladie** chez les exposés et les non exposés

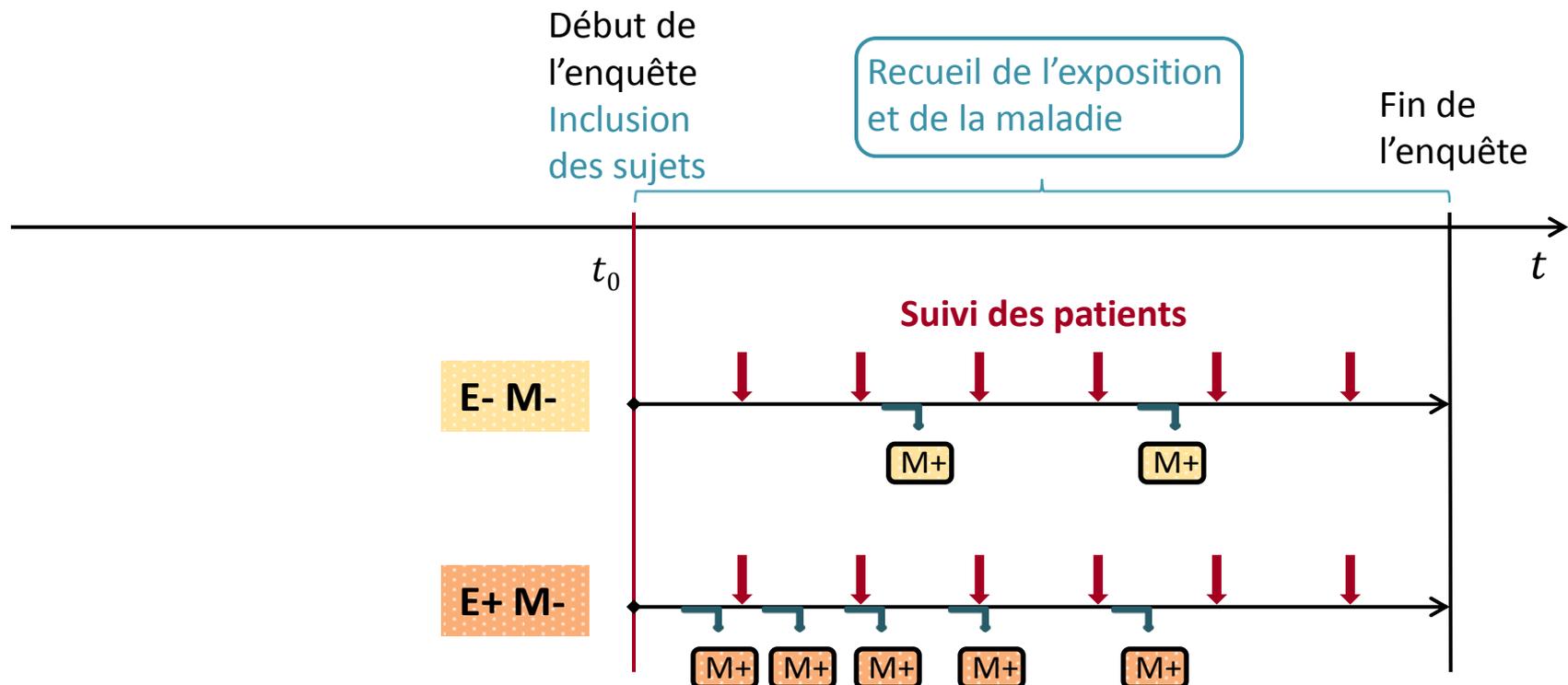
# III. Méthodologie générale

---

- Type de cohorte selon la **chronologie du recueil de données**
  - ↳ Cohorte prospective, cohorte rétrospective, cohorte historico-prospective
- Type de cohorte selon de **recrutement des sujets**
  - ↳ Cohorte fixe, cohorte dynamique, cohorte mixte
- Type de cohorte selon le **choix du groupe de référence**
  - ↳ Groupe de référence interne, groupe de référence externe
- Type de cohorte selon la **prévalence de l'exposition**
  - ↳ Prévalence forte et diffuse, prévalence faible et concentrée

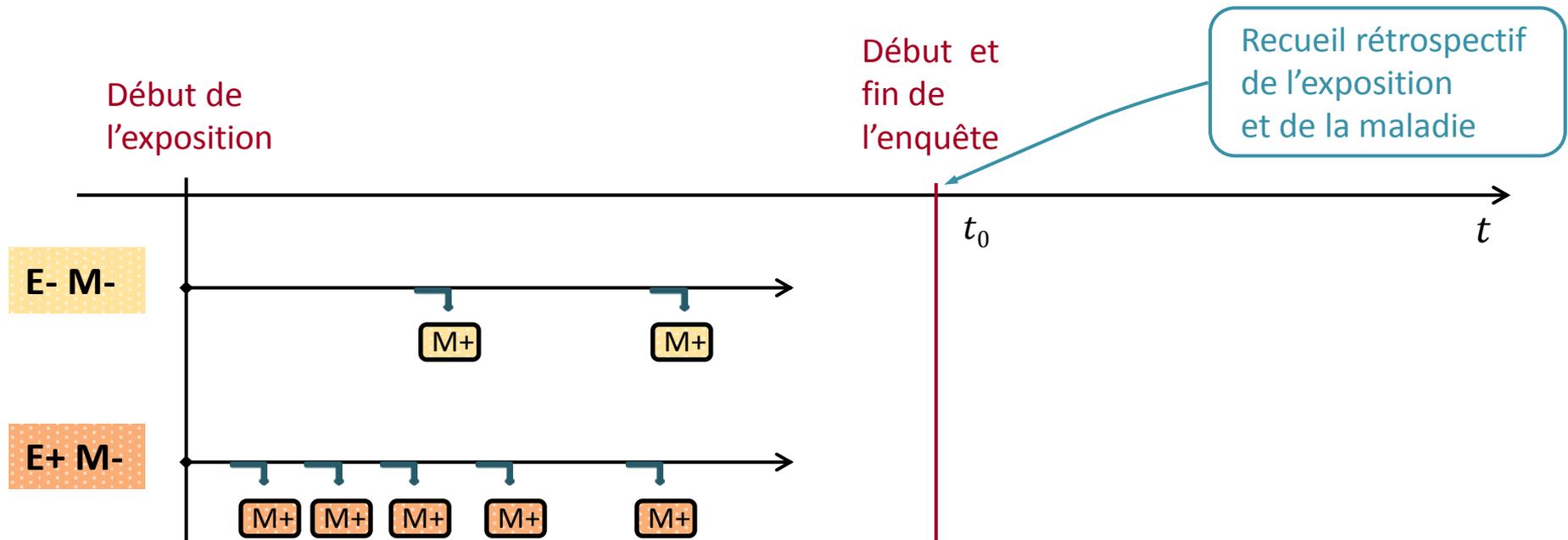
# Cohorte prospective

- Début de l'enquête au moment de l'exposition
- **Recueil prospectif** de l'exposition et de la maladie
- **Suivi des patients** (identique) dans les groupes E+ et E-



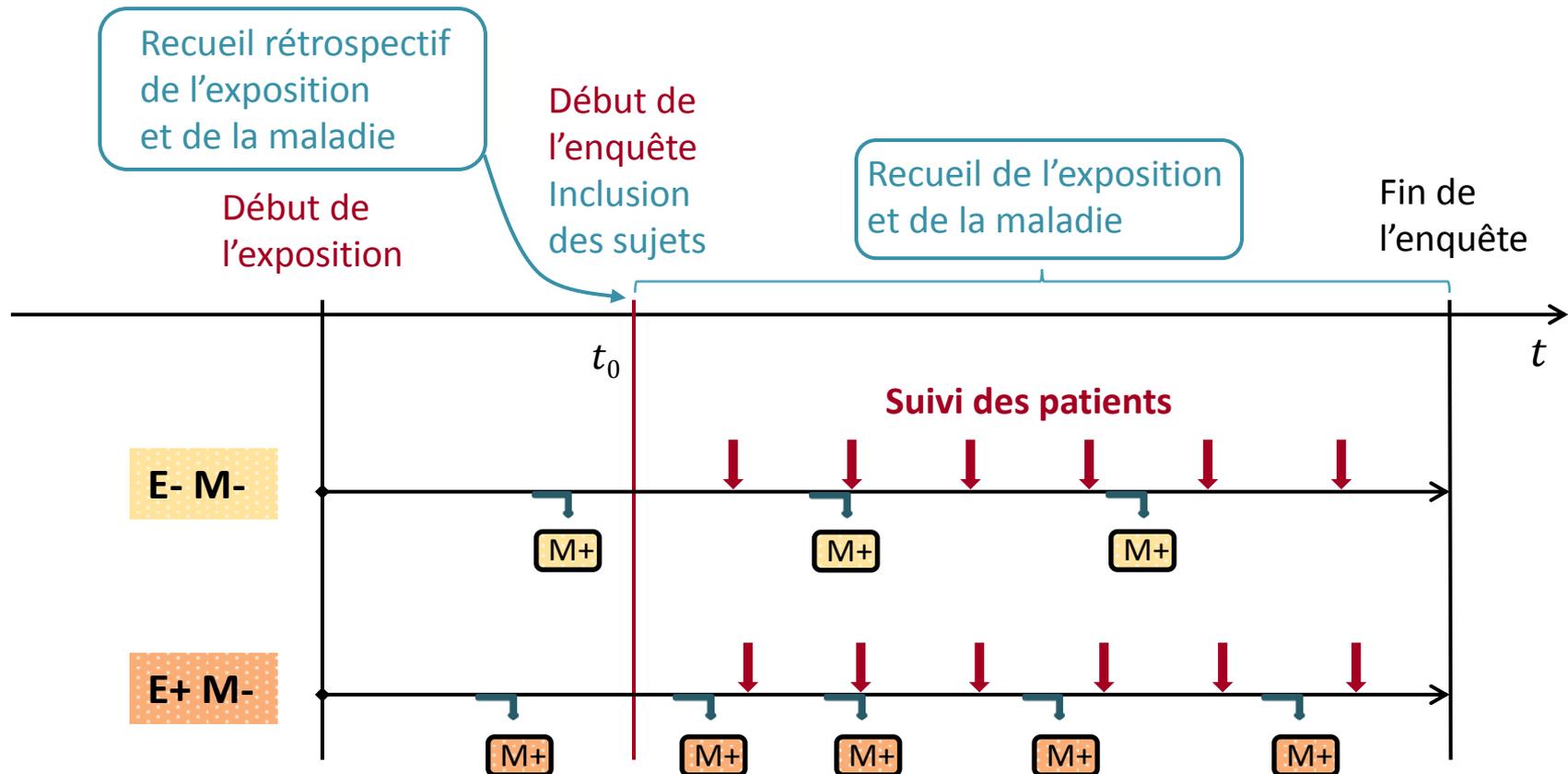
# Cohorte rétrospective

- Début de l'enquête après l'exposition (Ex: effet du Mediator à partir des fichiers de la sécurité sociale)
- **Reconstitution a posteriori de la cohorte** (reconstitution rétrospective d'un suivi individuel à partir d'un fichier disponible)
- **Recueil rétrospectif** de l'exposition et de la maladie



# Cohorte historico-prospective

- **Recueil rétrospectif et prospectif** de l'exposition et de la maladie
- Ex : inclusion des patientes au 3<sup>ème</sup> mois de grossesse dans l'étude du lien entre pathologie de la grossesse (E) et malformation de l'enfant (M)



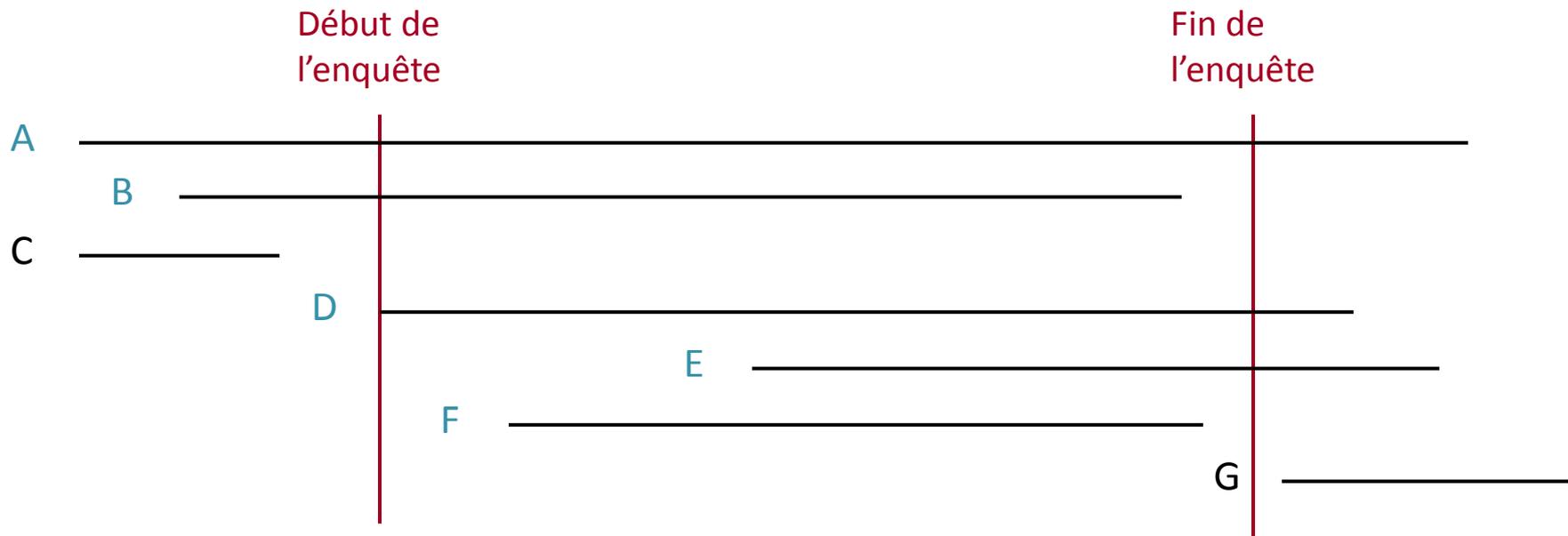
# Cohorte prospective ou rétrospective ?

---

- **Cohorte prospective**
  - (+) Recueil planifié des expositions et de la maladie
  - (+) Evaluation des expositions avant la survenue de la maladie
  - (+) Limite les perdus de vue
  - (-) Enquête longue et réponse retardée sur une exposition actuelle (temps de latence)
- **Cohorte rétrospective**
  - (+) Enquête et réponse rapide sur un risque lié à exposition passée
  - (-) Recueil hétérogène et souvent incomplet des expositions et des maladies
  - (-) Nombreux perdus de vue
  - (-) Information sur les facteurs de confusions potentiels rare et peu fiable

# Mode de recrutement des sujets

---



- **Cohorte fixe** : les sujets présents le jour ou pendant la période d'inclusion participent à l'étude (A,B,D)
- **Cohorte dynamique** : tous les nouveaux sujets présentant les critères d'inclusion depuis la date de début jusqu'à la date de fin de l'enquête (D, E, F)
- **Cohorte mixte** : les sujets présents et nouveaux participent à l'enquête (A, B, D, E, F)

# Mode de recrutement des sujets

---

- **Cohorte fixe**
  - **Biais de sélection** possible particulièrement en milieu de travail : si l'exposition est liée à une maladie silencieuse, les sujets sélectionnés sont les plus résistants et les moins exposés (sujets prévalents) : «**Healthy worker effect**»
  - Recrutement **rapide**
- **Cohorte dynamique**
  - Limite le **biais de sélection** (sujets incidents)
  - Recrutement **long**
- **Cohorte mixte**
  - Permet **d'étudier l'effet du biais de sélection** en comparant les sujets incidents et prévalents

# Choix du groupe de référence

---

- Définition du groupe de référence
  - Sujets non malade et non exposé aux facteurs étudiés
  - Suivi et recueil de l'information sur la maladie possible et similaire au groupe exposé
  - Sujets comparables au groupe exposé sauf sur l'exposition
    - Issu de la même population source que le groupe exposé
    - Idéalement, comparable sur toutes les caractéristiques ayant un lien avec la maladie étudiée (âge, sexe, catégorie socioprofessionnelle, lieu d'habitation, ...)
    - Les différences peuvent constituer des facteurs de confusion potentiels

# Choix du groupe de référence

---

- **Groupe de référence interne**
  - **Le plus utilisé** dans les études prospectives
  - Les groupes **E+ et E-** sont définis **au sein de la même cohorte** (avant ou après le recrutement)
  - **Suivi identique** des groupes E+ et E- (mesures standardisées)
  - Mesure de **l'exposition connu pour chaque individu** (idéalement, suffisamment élevée et hétérogène pour étudier une relation dose-effet)
  - **Ajustement possible** sur des facteurs de confusions individuels (variables mesurées à l'inclusion et pendant le suivi)

# Choix du groupe de référence

---

- **Groupe de référence externe**
  - Le groupe E- n'est pas issu de la **cohorte** (population nationale, registre, ...)
  - Nécessite **une population de référence** où l'incidence de la maladie est connue
  - **Utilisé dans les enquêtes rétrospectives** (commode quand on dispose déjà de statistiques sur une population)
  - **Rarement dans les études prospectives** (sauf si l'exposition est homogène ou non connu individuellement dans la cohorte)
- **Inconvénients :**
  - **Pas de mesures des facteurs de confusion** : ajustement limité à quelques facteurs socio-démographiques (âge, sexe, catégorie socioprofessionnelle, lieu d'habitation, ...)
  - **Nombreux biais** ➔ plusieurs références externes !
    - population source différente (biais de sélection)
    - Si prévalence de l'exposition importante dans la population générale (biais de classement)

# Prévalence de l'exposition

---

- Prévalence de l'exposition élevée et diffuse
  - Enquête sur un échantillon de population générale avec groupe de référence interne
  - Ex : pollution atmosphérique, facteurs alimentaires, tabac, alcool, consommation de médicaments courants,...
- Prévalence de l'exposition faible et concentrée
  - Cohorte exposée issue d'un groupe de sujets particulièrement exposé (Ex: individus d'un groupe professionnel)
  - groupe de référence interne ou externe (souvent la population générale)
  - Ex : expositions professionnelles

# IV. Mise en place d'une enquête exposé-non exposé

---

- **Constitution de la cohorte**
  - Identification de la population source
  - Critères d'inclusion
- **Suivi et recueil de l'information**
  - Modalités de suivi
  - Mesure de l'exposition
  - Mesure des critères de jugements
  - Mesure des facteurs de confusion
- **Suivi, perdus de vue et biais dans une cohorte**

# Constitution de la cohorte : population source

---

- **Critères pour choisir la population source**
  - Facilité de recueil des informations (ex : cohorte captive, médecine du travail)
  - Motivation des sujets (ex : médecins, infirmières, adhérents d'une mutuelle)
- **Exemples de population source**
  - Population générale (grandes bases de données nationales)
  - Patients consultants dans un ou plusieurs centre(s) hospitalier(s)
  - Adhérents à une assurance, une mutuelle (Ex: MGEN), un ordre professionnel (Ex : médecins britanniques pour l'étude tabac et cancer du poumon, Doll et Hill 1951-1954)
- Si l'exposition est rare, la population source devra être importante pour permettre le recrutement de suffisamment de sujets exposés

# Constitution de la cohorte : critères d'inclusion

---

- Sujets non atteints par la maladie étudiée à l'inclusion
- Sujets pour lesquels on dispose d'une information précise sur l'exposition
  - Plusieurs niveaux d'exposition pour l'étude d'une relation dose-effet
  - Durée d'exposition suffisante pour la survenue de la maladie
- Sujets pour lesquels la durée de suivi est suffisante
  - Eviter les perdus de vue (accès au statut vital indispensable)
  - Exclusion de certains individus (étrangers, les sujets difficiles à suivre (travailleurs intérimaires), ...)
- Minimiser la non participation à l'enquête (biais de sélection possible) en informant et en expliquant la problématique aux individus

# Constitution de la cohorte : biais

---

- **Biais de sélection** liés au mode de recrutement dans une **cohorte fixe (cas prévalents)**
  - Particulièrement en milieu professionnel où on sélectionne les sujets les plus résistants et les moins exposés (Healthy worker effect)
- **Biais de sélection** lors de la constitution de la cohorte :
  - l'échantillon de l'enquête n'est **pas représentatif** de la population cible\*
  - le groupe exposé n'est **pas comparable** au groupe non exposé\*
- Biais liés au **choix du groupe de référence externe**
  - **Biais de sélection** si le groupe exposé pas comparable à la population de référence\*
  - **Biais de classement** si la prévalence de l'exposition est élevée dans la population de référence

\* pour les caractéristiques ayant un lien avec la maladie étudiée

# Suivi et recueil de l'information

---

- **Recueil de l'information au cours du suivi**
  - Variables sociodémographiques (sexe, date de naissance, ...),
  - Exposition
  - Statut vital ou statut de la maladie
  - Facteurs de confusion potentiels
  - Dates de recrutement, date des dernières nouvelles, perdus de vue
- **Modalités de suivi**
  - Suivi identique dans les deux groupes (guides de procédure)
  - Durée du suivi dépend du délai d'apparition de la maladie (long pour certains cancers)
  - Suivi programmé : convocation, relance, enquêteur ou questionnaire postal
  - Veiller à la fiabilité du recueil (formation des enquêteurs, contrôle qualité)
  - Minimiser les données manquantes et les perdus de vue (explications, informations, contacts)

# Recueil de l'exposition

---

- Les études de **cohorte** permettent la **mesure précise de l'exposition** (absence de biais de mémorisation des enquêtes cas-témoin)
- **Recueil de l'exposition** doit être
  - Mesuré avec un **outil validé, fiable et standardisé** (questionnaire, interrogatoire, appareil de mesure, examen clinique, examens biologiques, prélèvements environnementaux ...)
  - Détaillé avec **différentes intensité** dans les niveaux d'exposition
  - Complet avec la **date de début, la durée, les changements d'intensité**

# Recueil du critère de jugement

---

- Critère de jugement : **mortalité, morbidité, qualité de vie**
- **Recueil du critère de jugement** doit être
  - Mesuré avec un **outil validé et fiable** (examens biologiques, interview avec un questionnaire validé, CépiDC INSERM)
  - **Standardisé et identique** pour les exposés et non exposés
  - **Effectué à l'aveugle** du groupe d'exposition (si pas d'outil objectif)
  - **Daté pendant le suivi** pour vérifier la chronologie E-M, estimer la durée d'apparition de la maladie et étudier l'évolution de la maladie

# Recueil du critère de jugement

---

## Mortalité

- (+) standardisé, accessible, exhaustif
- (+) diagnostic indépendant des expositions
- (+) valeurs de référence connues en population générale
- (-) certaines causes mal classées
- (-) non adapté aux pathologies non létales
- (-) ne permet pas toujours d'évaluer le risque de survenue de la maladie ou la qualité du système de soin

## Morbidité (état de santé)

- (+) interprétation plus facile pour évaluer le risque de survenue d'une maladie
- (-) moins standardisé, moins accessible et valeurs de référence pas toujours connues en population générale
- (-) Risque de biais différentiels selon l'exposition

## Qualité de vie

- Dépend des thématiques
- Existence de questionnaires validés

# Suivi, perdus de vue et biais

---

- Le **suivi de patients** dans le temps engendre des individus **perdus de vue**
- **Perdus de vue** : personnes dont le **suivi est interrompu** pour différentes raisons (déménagement, changement de médecin, état de santé qui s'améliore, ...)
  - ➔ état de santé **non déterminé** à la fin de l'étude
- **Biais de sélection** si le fait d'être **perdu de vue est lié à l'exposition** ou à la maladie
  - Ex : cohorte professionnelle avec une mauvaise tolérance à l'exposition
    - ➔ les individus fortement exposés sont plus souvent perdus de vue (pré-retraite, reclassement, licenciement, ...)
- **Limiter le nombre de perdus de vue** :
  - % de perdus de vue : critère de qualité d'une étude de cohorte
  - Suivi exhaustif des exposés et des non-exposés jusqu'à la fin de l'étude
  - Recherches complémentaires pour connaître le devenir des perdus de vue

# V. Avantages et limites des études de cohorte

---

- **Avantages**
  - Adaptés quand l'exposition est rare
  - Permettent d'examiner les conséquences multiples d'une même exposition
  - Permettent d'établir la séquence chronologique exposition-maladie
  - Recueil prospectif limite les biais de mesure de l'exposition et de la maladie
  - Estimation de l'incidence de la maladie dans les groupes exposé et non-exposé
  - Niveau de preuve supérieur à celui des enquêtes cas-témoins

# Avantages et limites des études de cohorte

---

- **Limites**
  - Non adaptée aux maladies rares ou à un long délai d'apparition
  - Prospectives : longues et coûteuses
  - Rétrospectives : biais possibles, nécessitent de disposer des informations appropriés
  - Validité des résultats à discuter en fonction des difficultés du suivi : perdus de vue, changement des expositions au cours du temps, modification des critères diagnostiques

# Pour approfondir !

---



## ***STROBE Statement***

Strengthening the reporting of observational studies in epidemiology

---

**STROBE** stands for an international, collaborative initiative of epidemiologists, methodologists, statisticians, researchers and journal editors involved in the conduct and dissemination of observational studies, with the common aim of Strengthening the Reporting of OBservational studies in Epidemiology.

<http://www.strobe-statement.org/>

<http://www.strobe-statement.org/index.php?id=available-checklists>

# Epidémiologie

## 3. Enquête cas-témoins

**Philippe SAINT PIERRE**

Université Paul Sabatier – Toulouse III

Institut de Mathématiques de Toulouse

[philippe.saint-pierre@math.univ-toulouse.fr](mailto:philippe.saint-pierre@math.univ-toulouse.fr)

# 3. Enquête cas-témoins

---

## I. Définition d'une enquête cas-témoins

## II. Méthodologie générale

- Recueil de données, mode de recrutement des sujets , choix des témoins

## III. Recueil des informations

- Réalisation pratique
- Mesure de l'exposition et de la maladie

## IV. Avantages et limites

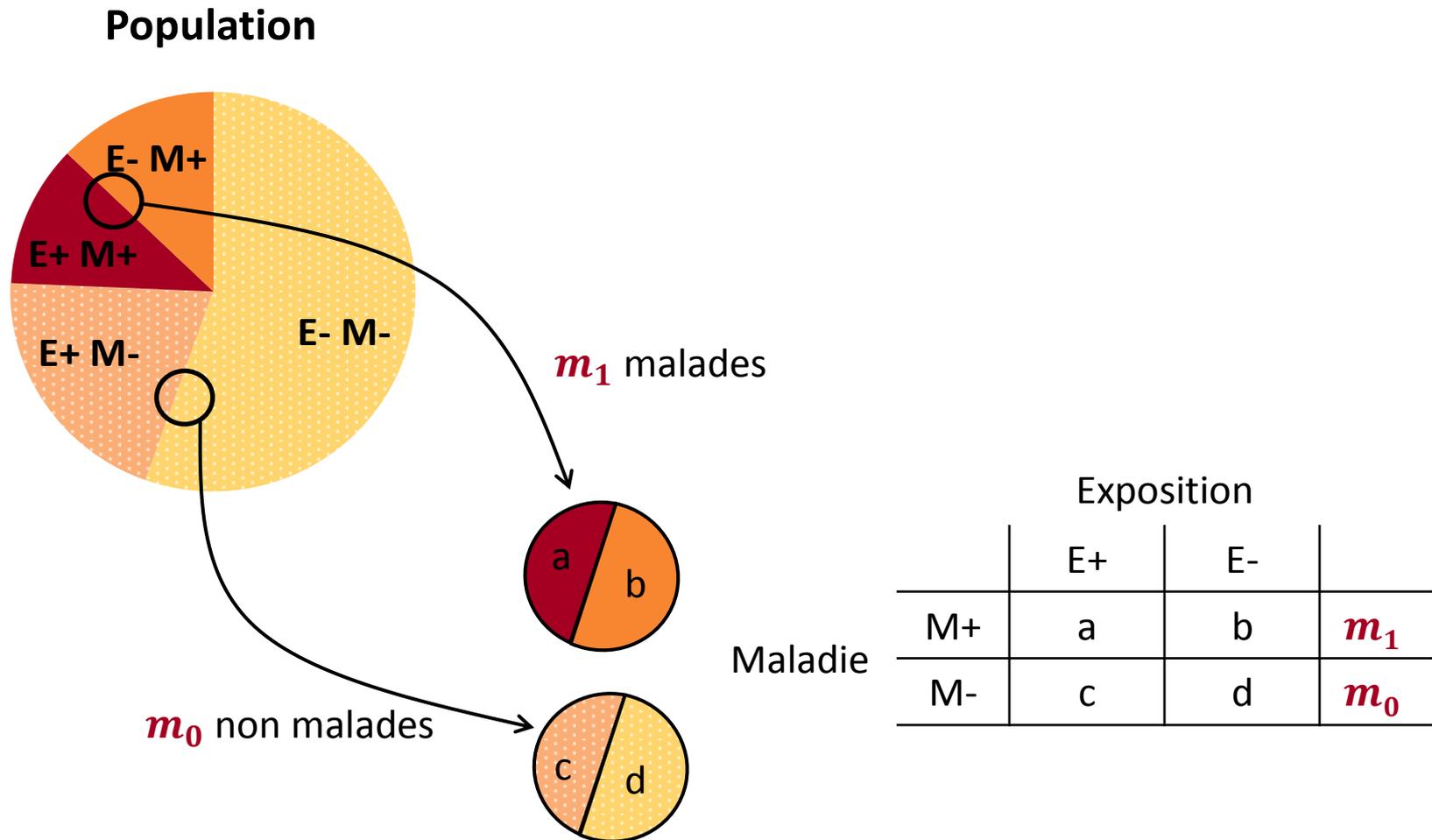
# I. Définition d'une enquête cas-témoins

---

- Objectif : comparer la répartition (fréquence) d'une exposition entre deux groupes appelés "cas" et "témoins"
- Le facteur à expliquer appelé "maladie"
  - Etat de santé, évènement de santé, pathologie, ...
- Constitution de deux groupes
  - Cas qui présentent l'évènement de santé
  - Témoins qui ne présentent pas l'évènement de santé
- Mesure de "l'exposition" à un facteur donné antérieure à la maladie
- Plusieurs autres facteurs explicatifs possibles
  - Facteurs de risque
  - Facteurs protecteurs

# Définition d'une enquête cas-témoins

- Sélection d'un échantillon de  $m_1$  malades et de  $m_0$  non-malades



# Exemple d'enquêtes cas-témoins

---

- **Question précise ou hypothèse à tester**
  - **Ex :** les femmes atteintes de cancer du sein ont-elles plus antécédents de prise d'oestroprogestatifs que les autres (indemnes de cette maladie)?
  - **Ex :** les enfants leucémiques ont-ils plus souvent que d'autres (non leucémiques) une mère qui a reçu une irradiation pendant sa grossesse ?
- **Etude exploratoire**
  - Phénomène de sante important non expliqué (urgence)
  - Pas d'idée sur la cause possible de la maladie
  - **Ex :** Nombre de cas de cancer du vagin anormalement élevé chez des adolescentes dans le Massachussetts en 1971
    - ➔ Enquête exploratoire sur 8 cas et 32 témoins
    - Prise de Distilbène par la mère chez 7 cas sur 8 et jamais chez les témoins

## II. Méthodologie générale

---

- **Recueil de l'information**  **toujours rétrospectif !**
  - Recherche de l'exposition effectué au moment de l'inclusion dans l'étude pour les cas et les témoins
  - Possible uniquement sur les patients vivant au moment de l'étude ou si l'information peut être donnée par l'entourage ou par un dossier médical
  - Risque de biais de classement (différentiel ou non)

# Mode de recrutement

---

- Inclusion des cas (malades)
  - Cas prévalents : nouveaux et anciens malades accessibles au moment de l'inclusion
    - Ne tiens pas compte de l'ancienneté du diagnostic
    - Risque de biais de sélection : survie sélective
      - ↳ prise en compte des "vieux" cas : plus résistants et moins exposés
    - Biais de classement (si modification de l'exposition et des comportements entre le diagnostic de la maladie et le début de l'étude)
  - Cas incidents : uniquement les nouveaux malades
    - Contrôle de la définition et du diagnostic de la maladie
    - Limite le biais de survie sélective et le biais de classement
    - Période de recrutement plus longue (nécessite un tau d'incidence élevé)
    - Plus de sites participants (hétérogénéité possible)
- Inclusion des témoins (non malades) ➡ Issus de la même population que les cas

# Choix des témoins

---

- **Témoins représentatifs de la population dont sont issus les cas** (comparable aux cas en tous points sauf pour l'exposition et la maladie)
  - Effectifs **équilibré** (1 cas pour 1 témoin) ou **déséquilibré** (1 cas pour 2 ou 4 témoins)
- 

Origine des cas

Sélection des témoins

---

**Population générale**

(Registres, dépistage, déclaration obligatoire)



**Population générale**

**Milieu hospitalier**

(hôpitaux, cliniques)



**Milieu hospitalier**

(même hôpital)

**Malades d'une cohorte**

(nested case control study)



**Non malade de la même cohorte**

---

# Sélection des témoins

---

- Témoins en population générale
  - Utilisation d'une base de sondage :
    - Liste de logement, listes électorales, certificats d'immatriculation, fichier de l'assurance maladie, appels au hasard de numéros de téléphone
- Témoins hospitaliers
  - Issus du même établissement que les cas
  - Hospitalisés pour une **pathologie sans lien avec la maladie ou l'exposition étudiées**
    - Exclure les maladies liés à l'exposition étudiée
    - Service où le recrutement est comparable à celui des cas (limiter les biais de sélection)
    - Considérer plusieurs services pour avoir plusieurs pathologie représentées (plusieurs groupes témoins possibles)

# Sélection des témoins

---

- Sélection des témoins peut être
  - **Aléatoire** : chaque sujet a une chance identique d'être tiré au sort
  - **Systematique** : tous les  $k^{\text{ièmes}}$  individus de la liste après tirage au sort du premier
  - **Apparié** : à chaque cas est associé par tirage au sort des témoins répondant à certains critères identiques (Ex : âge, sexe, CSP)
  - **Stratifié** : tirage au sort des témoins dans des sous groupes ou strates (Ex: région, ville) de la population cible (appariement par classe)



Appariement et stratification visent à équilibrer la répartition de certains facteurs de confusion dans le groupe des cas et des témoins

- Possible de **restreindre la population cible** pour avoir des cas et des témoins **suffisamment exposés** et homogènes (même âge, sexe, région, ...)

# Choix témoins : population générale ou hospitalier

---

## Témoins en population générale

## Témoins hospitalisés

|                                 |                               |                           |
|---------------------------------|-------------------------------|---------------------------|
| Accès aux témoins               | difficile                     | facilité                  |
| Motivation                      | faible                        | meilleure                 |
| Disponibilité                   | faible                        | meilleure                 |
| Taux de réponse                 | faible                        | meilleure                 |
| Zone géographique               | plus grande que celle des cas | comparable aux cas        |
| Biais de sélection              | plus faible                   | oui (admission sélective) |
| Biais de mémoire                | oui                           | plus faible               |
| Surexposition                   | peu probable                  | possible                  |
| Biais de mesure de l'exposition | Possible                      | possible                  |
| Généralisation des résultats    | plus large                    | restreinte                |

# Sélection des témoins : exemple de biais

- **Enquête cas- témoins appariée** : cancer de la vessie et consommation de cigarette

| Nombre de cigarettes | Odds ratio  |                |
|----------------------|---|----------------|
|                      | Paires dont le témoins à une maladie cardiovasculaire | Autres paires  |
| Non fumeur           | 1   | 1              |
| < 20                 | 1,48 (p > 0,05)                                       | 3,26 (p <0,01) |
| 20 – 39              | 1,28 (p > 0,05)                                       | 4,41 (p <0,01) |
| ≥ 40                 | 0,7 (p > 0,05)  | 6,92 (p <0,01) |

- Le tabac est associé aux maladies cardiovasculaires
  - ➔ L'appariement sur un témoin atteint d'une maladie cardiovasculaire revient à **appairer sur l'exposition**
- **L'association disparaît** par un mauvais choix du groupe témoins

# Sélection des témoins : exemple de biais

---

- **Enquête cas- témoins appariée** : facteurs de risque d'une grossesse extra-utérine (GEU)
  - **Cas** : femmes traitées pour une GEU dans plusieurs maternités (**M = 903**)
  - **Témoins** : femmes ayant accouchée juste après dans la même maternité (**N = 1527**)
  - **Pb** : certains témoins ont déjà eu des GEU (connaissent les facteurs de risques)
  - **Pb** : les témoins ont tous choisi de mener leur grossesse à terme (couple "stable")
- **Restriction** aux femmes vivant maritalement et sans contraception

$$M = 903 \quad \rightarrow \quad M' = 570 \quad (63\%)$$

$$N = 1527 \quad \rightarrow \quad N' = 1385 \quad (90,7\%)$$

# III. Recueil des informations

---

- **Qualités d'un outil de recueil**
  - **Validité** : capacité à mesurer ce qu'il est censé mesurer
  - **Reproductibilité** : les mêmes réponses doivent être obtenues quand on soumet le questionnaire plusieurs fois à la même personne dans les mêmes circonstances
  - **Concordance** : les mêmes réponses doivent être obtenues quand deux enquêteurs différents soumettent le questionnaire à la même personne
  - **Acceptabilité** par le patient et **facilité** de réalisation du recueil
- **Informations, explications et relance** des sujets pour éviter la non participation et les données manquantes

# Mesure de l'exposition

---

- Toujours **rétrospectif** et obtenue à partir **d'archives** (dossiers médicaux...), interview (direct ou téléphonique), auto-questionnaire postal
- Risque de **biais de classement** (notamment différentielle) liées au recueil d'informations anciennes (**biais de mémorisation**)
- Pour **éviter les biais de classement**
  - Utiliser des **informations objectives** (marqueurs biologiques, Ex : l'HbA1c pour l'exposition au diabète)
  - **Formation** des enquêteurs
  - **Standardisation et contrôle** de l'outil de recueil
  - Recueil "**à l'aveugle**" du statut cas ou témoins
  - Recueil dans les **mêmes conditions** pour les cas et les témoins
  - Attribuer des cas et des témoins à chaque enquêteur

# Mesure de la maladie

---

- Critères et méthodes de diagnostics bien définis et standardisés
- Réalisation identique et dans les mêmes conditions pour tous les sujets (cas et témoins)
- Si possible "à l'aveugle" du statut exposé ou non exposé
- Sinon éviter que le diagnostic soit fait à l'occasion d'un examen motivé par une exposition particulière

# IV. Avantages et limites des études cas-témoins

---

- **Avantages**

- Adaptées aux **maladies rares** et à un long délai d'apparition
- Souvent plus courtes et moins coûteuses que les cohortes

- **Limites**

- Pas adaptées aux **exposition rares**
- **Choix des témoins** souvent difficile (biais de sélection)
- **Recueil rétrospectif** de l'exposition (biais de mémorisation)
- **Pas d'estimation de la prévalence** (nombre de cas fixé), de l'incidence ou du risque de la maladie (pas de suivi)
- **Niveau de preuve inférieur** à celui des enquêtes de cohorte

# Pour approfondir !

---



## ***STROBE Statement***

Strengthening the reporting of observational studies in epidemiology

---

**STROBE** stands for an international, collaborative initiative of epidemiologists, methodologists, statisticians, researchers and journal editors involved in the conduct and dissemination of observational studies, with the common aim of Strengthening the Reporting of Observational studies in Epidemiology.

<http://www.strobe-statement.org/>

<http://www.strobe-statement.org/index.php?id=available-checklists>