

### 3.6. Les sources d'erreurs lors d'une enquête épidémiologique

Les études épidémiologiques comportent de nombreux risques d'erreurs. Les épidémiologistes consacrent beaucoup d'attention à la réduction de ces erreurs et à l'évaluation de l'effet de celles qui ne peuvent être éliminées. Les erreurs peuvent être aléatoires ou systématiques.

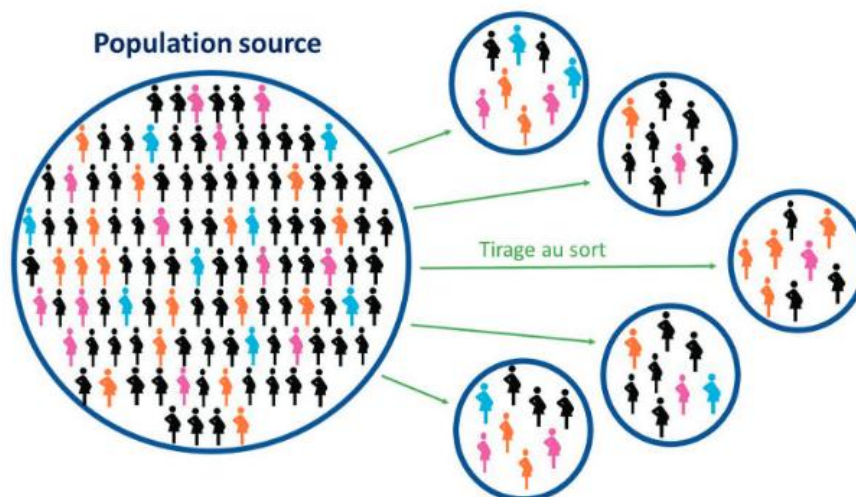
#### Erreurs aléatoires et systématiques

On distingue deux grands types d'erreurs :

- les erreurs dites aléatoires, qui sont des erreurs liées aux fluctuations d'échantillonnage ;
- les erreurs systématiques ou biais, qui sont des erreurs de conception de l'étude.

#### 3.6.1. Erreur aléatoire

Constituer un **échantillon de répondants** revient à créer un raccourci pour mener son enquête. Mais un échantillon dont le recrutement a lieu au sein de la population source peut néanmoins différer de cette dernière, comme dans la **figure 1** où les cinq échantillons successivement constitués sont composés de sujets ayant des caractéristiques différentes, ce qui est complètement lié au hasard. Il s'agit d'une erreur aléatoire, également appelée fluctuation d'échantillonnage.



Ce type d'erreur ne remet toutefois pas en cause les résultats puisque les mesures d'association calculées dans chacun des échantillons vont se situer autour de la valeur qui aurait été obtenue si les données de toute la population source avaient été disponibles .

## ❖ Choisir la bonne taille pour son échantillon

Pour obtenir des résultats fiables, la première chose à faire est de **bien déterminer la taille de votre échantillon**. Et ce travail est à renouveler pour chaque nouvelle recherche car en fonction de la typologie de votre population cible ou la nature des résultats attendus vous n'allez pas interroger le même nombre de personnes.

### A savoir :

- Taille de la population de référence (ou du moins, une bonne approximation)
- Le niveau de confiance attendu
- La marge d'erreur acceptable (également appelée **intervalle de confiance**)

**DONC :** Une erreur aléatoire consiste dans l'écart (du fait du seul hasard) entre une observation faite dans un échantillon et la valeur vraie dans la population, d'où un manque de précision dans la mesure d'une association.

### Il existe trois sources principales d'erreur aléatoire:

- les variations biologiques individuelles ;
- les erreurs d'échantillonnage ; et
- les erreurs de mesure.

La meilleure façon de réduire les erreurs d'échantillonnage consiste à augmenter la taille de l'échantillon. Elle doit être suffisamment importante pour que l'étude ait la puissance statistique voulue pour détecter les différences considérées comme importantes.

### La détermination de la taille d'un échantillon nécessite qu'on soit renseigné sur les paramètres suivants :

- Capacité à déceler une différence avec le degré de signification statistique requis;
- Erreur ou probabilité jugée acceptable pour qu'un effet réel passe inaperçu ;
- Importance de l'effet étudié ;
- Fréquence de la maladie dans la population ;
- Tailles relatives des groupes comparés

### 3.6.2. Erreurs systématiques

Les **erreurs systématiques** ou les biais doivent, dans la mesure du possible, être anticipés, notamment grâce à la rédaction d'un protocole extrêmement précis et objectif

les erreurs dites systématiques, ou biais, qui éloignent dans une direction précise les estimations calculées sur l'échantillon de la vraie valeur dans la population source (figure 2).

Dans une étude où l'erreur systématique est faible, on dit que l'exactitude est grande. L'exactitude ne dépend pas de la taille de l'échantillon.

Dans les études épidémiologiques, trois sortes de biais peuvent être rencontrées : de sélection, de classement et de confusion. Ils doivent être prévenus en amont pour que leurs conséquences sur les résultats soient limitées .

Les deux premiers types de biais (sélection et classement) ne peuvent être évités qu'au moment de la réalisation de l'étude, alors que le biais de confusion peut être pris en compte lors de la planification du protocole, ainsi qu'au moment de l'analyse statistique.

### ❖ Définition du biais

C'est une erreur systématique qui entraîne une déviation de ce qui est observé par rapport à la réalité. Si on étudie l'association entre un facteur de risque et une maladie donnée, l'erreur est dans la même direction à chaque fois. Cela peut donc entraîner une surestimation ou une sous-estimation d'un effet.

### A-Biais de sélection

On regroupe sous ce terme tous les biais qui conduisent à ce que les sujets inclus dans l'enquête ne soient pas représentatifs de la population étudiée. On doit particulièrement être vigilant au biais de recrutement. Ce dernier apparaît chaque fois que la probabilité d'entrée d'un sujet dans l'étude est liée à certains facteurs étudiés.

Les biais de sélection correspondent aux erreurs commises lors de la constitution de l'échantillon d'enquête et de l'inclusion des sujets dans l'étude. Ils sont impossibles à corriger une fois que la sélection des participants est terminée. Il est tout de même nécessaire de les identifier pour tenter de déterminer de quelle manière ils sont susceptibles de biaiser les mesures d'association.

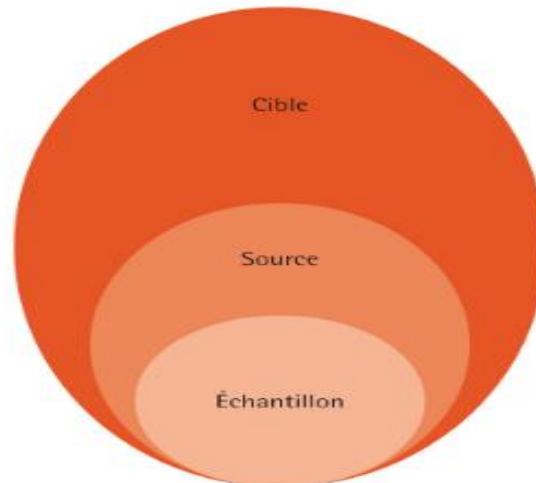
### Exemple de biais de sélection

- **Échantillon** : patients inclus dans une étude.
- **Population source** : population dont est tiré l'échantillon : par exemple, tous les patients vus dans un service pendant la période d'étude, tous les patients hospitalisés dans un service de neurologie pour un accident vasculaire cérébral (AVC) pendant la période d'étude, etc.
- **Population cible** : population à laquelle on veut généraliser les résultats : par exemple, tous les patients souffrant d'AVC en Algérie.

On parle de biais de sélection lorsque l'échantillon n'est pas représentatif de la population cible (3 raisons principales à cela) :

- **échantillon** non représentatif de la population source : le seul médecin qui inclut des patients dans une étude monocentrique est le spécialiste régional d'une forme rare de la maladie ;

- **population source** non représentative de la population cible : inclure des patients uniquement dans des centres experts qui présentent des pathologies avec des caractéristiques particulières et rares ;
- **sorties d'étude** : si l'échantillon est réduit par des sorties d'études (on parle d'attrition), l'échantillon restant est sélectionné et par conséquent moins représentatif de la population source et donc de la population cible.



## **B-Biais de classement ou de mesure**

**Un biais de classement**, également appelé biais d'information ou de mesure, correspond à une erreur dans le recueil et le classement des sujets dans leur statut d'exposés/non exposés ou de malades/non malades. Il ne peut pas être corrigé une fois l'étude terminée et les données recueillies. Cependant, il est intéressant de savoir à quel point il affecte les mesures d'association calculées. En outre, un biais de classement est analysé selon son caractère différentiel. Il convient alors de se demander si l'erreur affecte de la même manière les deux groupes de l'étude ou principalement l'un plutôt que l'autre .

### **L'erreur est dite différentielle lorsque :**

- l'erreur de mesure ou de classement sur l'exposition est influencée par la connaissance de la maladie :

- ce type d'erreur est impossible si l'étude est prospective (on recueille l'exposition alors que la maladie n'est pas encore survenue),

- dans une étude cas-témoins, on interroge des patients sur leur consommation de tabac après le diagnostic de cancer bronchique. Les cas ont tendance à surestimer leur consommation, tandis que les témoins la sous-estiment. Le lien entre le cancer et le tabac est donc artificiellement augmenté ;

- l'erreur de mesure ou de classement sur la maladie est influencée par la connaissance de l'exposition :

- ce type d'erreur est rare dans les études rétrospectives, puisqu'au moment où l'on recueille l'exposition, la maladie est déjà connue,

- par exemple, lors d'une étude exposés/non exposés sur le lien entre la consommation de café et le cancer du côlon, la connaissance du lien potentiel va conduire à un sur-dépistage du

cancer du côlon par coloscopie chez les grands consommateurs de café, et donc à une augmentation artificielle de l'association ;

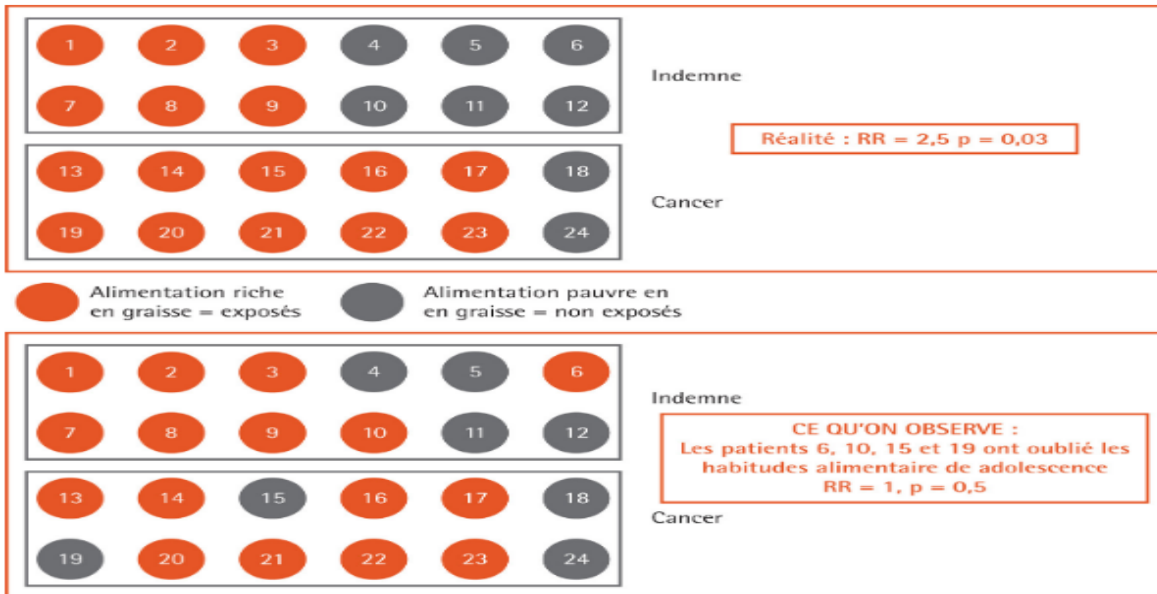
- l'évaluation de l'efficacité d'un traitement et des effets secondaires est faite en ouvert : par exemple, dans le cas d'un essai comparant l'efficacité d'un anti-biotique 7 jours versus 14 jours dans l'infection urinaire haute, si le critère de jugement est la récurrence des symptômes, et que l'essai est réalisé en ouvert, les patientes du groupe « 7 jours » signaleront plus facilement des symptômes après la fin du traitement que les patientes encore sous antibiotique. Ceci augmentera artificiellement l'efficacité du traitement de 14 jours.

**L'erreur est dite non différentielle** lorsque l'erreur affecte de manière aléatoire les deux groupes. Elles ont globalement des conséquences moins importantes sur les résultats que les erreurs différentielles.

Un des biais non différentiels peut advenir lorsque l'outil de mesure utilisé n'est pas fiable ou qu'il est inadapté à la recherche menée. Lorsque l'enquête est basée sur un questionnaire, il faut privilégier un outil testé et validé qui évaluera de manière fiable le phénomène étudié. Lorsqu'il s'agit de critères cliniques, un examen le plus objectif et standardisé possible doit être effectué pour recueillir les informations.

**Par exemple:** si l'objectif est de recueillir le poids de toutes les femmes enceintes incluses dans une étude épidémiologique lors d'une consultation au premier trimestre de la grossesse, mais que l'unique balance utilisée minimise le poids de 3 kg, le risque est de minorer la proportion de femmes obèses dans l'échantillon. Les mesures d'association calculées peuvent donc être sous-estimées. Pour limiter ce type de biais non différentiel, il faut vérifier la fiabilité des instruments utilisés.

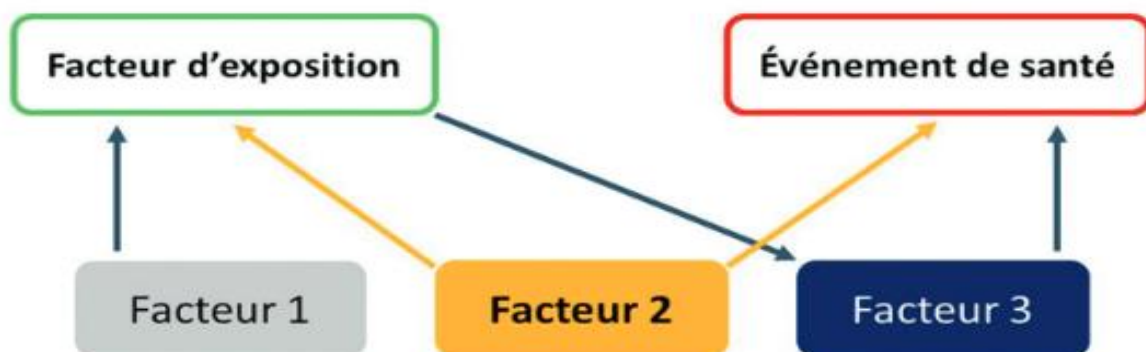
**Par exemple:** on interroge des sujets sains de 40 ans sur leurs habitudes alimentaires à l'adolescence, puis on les suit pendant 10 ans pour étudier le lien entre alimentation riche en graisse et cancer (étude historico-prospective). Le cancer n'étant pas survenu au moment de la mesure de l'exposition, une erreur différentielle est impossible. En revanche, certains patients ont oublié leurs habitudes alimentaires passées et répondront mal au questionnaire. Cette erreur affecte identiquement les deux groupes, et aura tendance à réduire l'estimation de l'association entre l'alimentation et le cancer (voir figure ci-dessous).



Pour éviter ces erreurs dans les enquêtes épidémiologiques, il est important de standardiser les modes de recueil et de suivi des deux groupes. De surcroît, lorsque les données sont collectées lors d'entretiens, les enquêteurs en charge de les réaliser peuvent être laissés en aveugle du groupe d'appartenance de la personne qu'ils interrogent. En effet, dans une étude cas-témoins dans laquelle un groupe de malades et un groupe de non-malades sont constitués, l'enquêteur peut être tenu à l'écart du statut de malade ou non malade de la personne qu'il interroge pour ne pas influencer sa manière de poser les questions.

### C-Biais de confusion

Il est question de biais de confusion lorsqu'un tiers facteur intervient dans la relation entre exposition et maladie



- Le **facteur 2** est un facteur **de confusion** car il est lié à l'exposition et à l'événement d'intérêt
- Le **facteur 1** est un facteur seulement lié à l'exposition
- Le **facteur 3** est un facteur **intermédiaire** puisque l'exposition est lié au facteur 3 lui-même lié à l'événement d'intérêt

On parle de biais de confusion lorsque l'association observée entre une exposition et une maladie est en réalité liée à un troisième facteur, associé à l'exposition (c'est-à-dire plus fréquent chez les exposés que chez les non-exposés) et lié à la maladie (figure 3).

**Par exemple :**

- On souhaite étudier le lien entre tabagisme et cirrhose. En analyse bivariée, la cirrhose est significativement plus fréquente chez les fumeurs. Or, la consommation d'alcool plus fréquente chez les fumeurs est fortement liée au risque de cirrhose. L'alcool est un facteur de confusion dans l'étude de la relation entre tabac et cirrhose ;

De manière générale, le fait de ne pas tenir compte de facteurs de confusion peut conduire à surestimer les associations mises en évidence. En revanche, cela n'aurait pas de conséquence lorsque aucune association n'est relevée.

**Conclusion:**

**Dans le tableau ci-dessous :**

- les fluctuations d'échantillonnage entraînent des erreurs aléatoires et donc une imprécision des résultats ;
- les biais entraînent des erreurs systématiques et donc une surestimation ou une sous-estimation des résultats.

