

## Causalité en épidémiologie

La notion de causalité est centrale en recherche épidémiologique puisqu'elle constitue l'une de ces finalités : déterminer les causes des maladies. Bien qu'il s'agisse en réalité d'un concept philosophique complexe, nous nous concentrons sur la notion de causalité dans le cadre de l'épidémiologie.

### Définir la causalité

La notion de causalité dans le domaine de la santé a beaucoup évolué au cours du temps. Dans les derniers siècles, cette évolution a suivi de près ce qu'on nomme la transition épidémiologique, c'est-à-dire le passage d'une mortalité due majoritairement aux maladies infectieuses à l'émergence des maladies et des expositions chroniques (diabète, maladies cardiovasculaire, cancers, etc.).

Ainsi, au 19<sup>e</sup> siècle, deux médecins allemands, Henle et Koch ont formulé de façon simple [quatre postulats](#) qui permettaient d'établir un lien de causalité entre un germe et une maladie :

1. Le micro-organisme doit être présent en abondance dans tous les organismes souffrant de la maladie, mais absent des organismes sains.
2. Le micro-organisme doit pouvoir être isolé de l'organisme malade et cultivé in vitro.
3. Le micro-organisme cultivé doit entraîner l'apparition de la maladie lorsque introduit dans un organisme sain.
4. Le micro-organisme doit être é nouveau isolé du nouvel organisme hôte rendu malade puis identifié comme étant identique à l'agent infectieux original.

Ces postulats très mécaniques et déterministes contrastent grandement avec les simples [arguments de causalité énoncés par Hill](#) un siècle plus tard. Notons d'ailleurs que si ces arguments sont souvent cités de nos jours comme les *critères de Bradford Hill*, leur auteur ne les a jamais présentés comme tel, mais bien comme des éléments de discussion en faveur de la causalité.

### L'argument contrefactuel

Afin de comprendre ce qui a causé cette complexification de la définition opérationnelle de la cause d'une maladie, revenons à une définition plus simple, bien qu'abstraite : « un facteur est la cause de la maladie d'un individu si en l'absence de ce facteur la maladie ne serait pas survenue. »

Cette définition, qui fait appel à ce qu'on appelle **l'argument contrefactuel**, est évidemment impossible à prouver dans la vie réelle. Mais c'est bien l'argument contrefactuel

qu'on cherche à approcher par l'appariement des individus ou la randomisation des expositions.

En utilisant cette approche pour rechercher les causes de différentes maladies, on peut se rendre compte qu'elles sont difficiles à déterminer de façon claire, notamment pour deux raisons principales : 1. les causes des maladies sont multiples, 2. ces causes sont souvent liées entre elles.

## Causes nécessaires et suffisantes

Prenons l'exemple du cancer du poumon : selon l'argument contrefactuel, le tabagisme peut être considéré comme une cause. Cependant, on sait que tous les fumeurs ne développeront pas un cancer du poumon et, à l'inverse que tous les cancers du poumon ne sont pas causés par le tabac. Le fait de fumer n'est donc une cause ni **suffisante** ni **nécessaire** du cancer du poumon. Cette situation est rencontrée bien souvent dans l'épidémiologie moderne : la combinaison d'un certain nombre d'exposition est nécessaire à la survenue d'une maladie. Ces combinaisons impliquent des facteurs biologiques, génétiques, comportementaux, environnementaux qui sont pour la plupart mal identifiés. **Et c'est parce qu'on ignore pour la plupart les combinaisons suffisantes au déclenchement des maladies qu'il est nécessaire d'utiliser la notion de risque.** On passe alors du cadre déterministe au cadre probabiliste. On ne parle dès lors plus de cause mais de **facteur de risque**.

## Réseaux de causes

En plus d'être multiples, les causes des maladies sont souvent intriquées et leurs effets propres difficiles à démêler. C'est ce qui cause les biais de confusion dont nous avons déjà parlé. L'ajustement, notamment via des modèles multivariés permet souvent de répondre à ce problème, mais le choix des variables d'ajustement est loin d'être évident. Plusieurs outils permettent d'aider dans ce choix, dont les **DAG**.

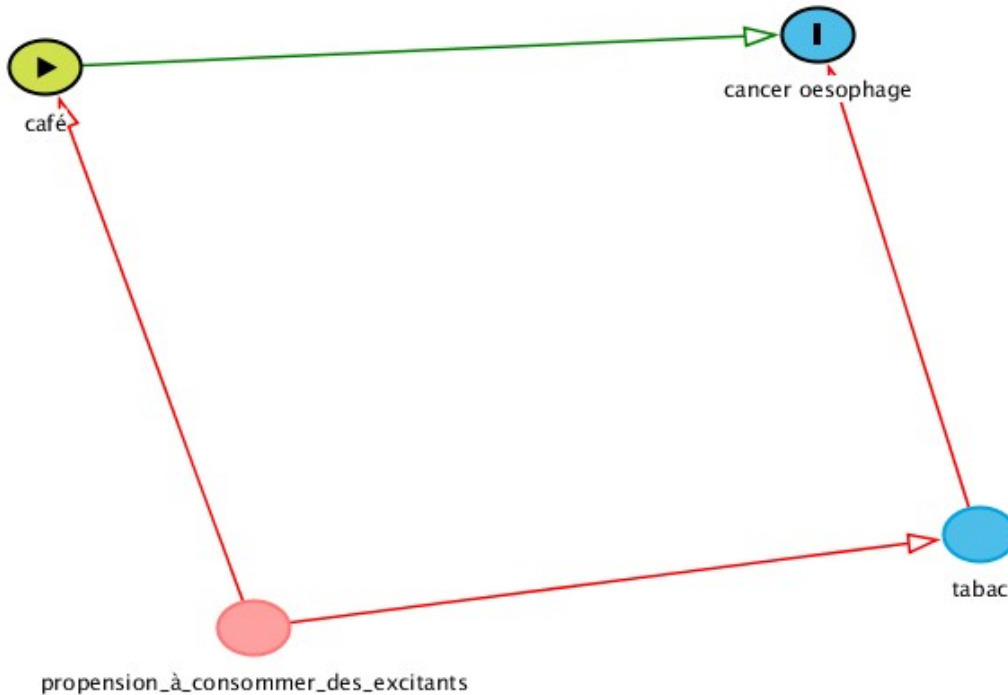
## les DAG

Un graphe orienté acyclique (GOA -- ou DAG en anglais) en épidémiologie correspond à :

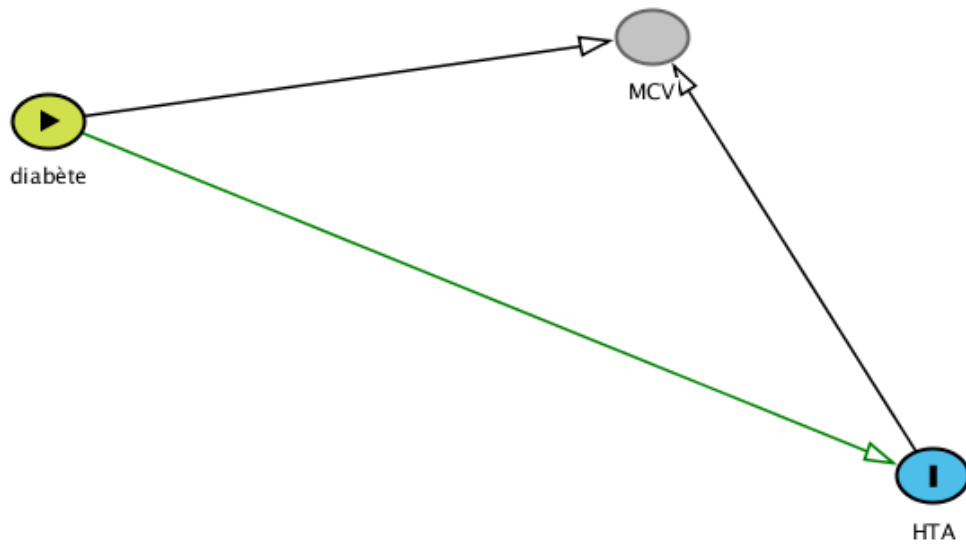
- **des noeuds** constitués par les expositions et maladies
- **des relations** entre certains de ces noeuds
- **un sens** causale dans les relations, matérialisé par une pointe de flèche

Les relations créent des **chemins** entre les différents facteurs qui peuvent être causaux ou détournés. Ainsi dans l'exemple ci-dessous, on s'interroge sur l'existence d'un chemin causal entre consommation de *café* et *cancer de l'œsophage*. Ce lien peut être biaisé en l'absence d'ajustement par le lien détourné passant par la *propension à l'utilisation d'excitants* et la consommation de *tabac* (notons au passage que les chemins détournés empruntent des relations sans tenir compte du sens de la flèche). C'est l'exemple classique de biais de confusion et l'ajustement sur le tabac sera nécessaire pour estimer correctement l'existence du lien causal *café/cancer*. L'ajustement permet ici de bloquer le chemin

détourné au niveau du noeud *tabac*. (En revanche, la propension à consommer des excitants est difficilement mesurable et donc ajustable ; on parle de variable latente).



Un chemin détourné est en revanche naturellement bloqué lorsqu'il contient une *collision*, qui survient quand deux des relations qui constituent ce chemin pointent directement sur le même nœud. Dans l'exemple ci-dessous, si nous nous intéressons au lien causal entre diabète et hypertension artérielle (HTA), il n'est pas nécessaire d'ajuster sur les maladies cardio-vasculaires, puisqu'il y a collision sur ce nœud. On peut même dire que cela aurait une conséquence fâcheuse puisque **l'ajustement sur une collision a pour effet d'ouvrir le chemin détourné**, et ainsi de causer un biais de confusion.



En suivant ces quelques principes il est aisé de choisir les variables d'ajustement pour des relations simples. En revanche, lorsque les relations se complexifient, comme c'est souvent le cas en épidémiologie environnementale ou sociale, par exemple, il est nécessaire d'utiliser des outils informatiques tels que [dagitty](#).