

**(Introduction à l') Introduction aux diagrammes causaux
(Directed Acyclic Graphs)**

Loïc Desquilbet

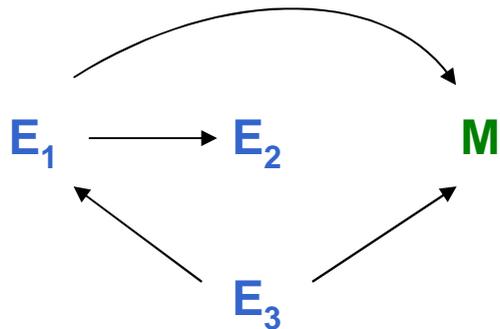
Contexte et représentation des diagrammes causaux

- Contexte

Recherche de facteurs de risque d'une maladie

⇔ Approcher au maximum la relation de causalité entre une exposition et une maladie

- Représentation générale



Où E₁, E₂, et E₃ sont des expositions, et M la maladie

Objectifs des diagrammes causaux

- Aider l'épidémiologiste à résumer **visuellement** les hypothétiques relations entre les différentes expositions et la maladie (« variables »)
- Obliger l'épidémiologiste à se poser la question de la temporalité des relations entre les variables
- Repérer et éviter les biais de classement, confusion, et sélection
- Permettre d'identifier de « faux » facteurs de confusion avant de construire et de « faire tourner » un modèle de régression

Rappel sur les biais de confusion (1)

- Contexte
 - Soit E une exposition d'intérêt principal, X une autre exposition, et M la maladie
 - On cherche à quantifier la relation entre E et M
- Définition d'un biais de confusion

L'association entre E et M est biaisée par du biais de confusion si E et M partagent une cause commune

Rappel sur les biais de confusion (2)

- Définition d'un facteur de confusion

X est un facteur de confusion si la non prise en compte de X dans les analyses statistiques introduit un biais dans la relation entre E et M

(X peut aussi être appelée « tiers facteur »)

- Prise en compte des facteurs de confusion dans les analyses statistiques
 - Standardisation
 - Sélection
 - Stratification
 - Ajustement dans un modèle multivarié

Rappel sur les biais de confusion (3)

- Exemple (1)
 - Un étudiant un peu naïf veut étudier la relation entre le risque de fracture de la hanche (M) et la présence de cheveux gris (E)
 - On suppose (mais l'étudiant ne le sait pas) qu'il n'y a aucune relation causale entre les cheveux gris et la fracture
 - L'étudiant trouvera une association brute entre le fait d'avoir des cheveux gris et celui d'avoir une fracture de la hanche :

$$\Pr(\text{cheveux gris} / \text{fracture}) > \Pr(\text{cheveux gris} / \text{pas de fracture})$$

Rappel sur les biais de confusion (4)

- Exemple (2)
 - L'association brute entre « cheveux gris » et « fracture » est biaisée (confusion) car ces variables possèdent une cause commune (ici, l'âge)
 - A âge égal, il n'y a plus d'association (statistique) entre cheveux gris et fracture
 - La prise en compte de l'âge a permis d'éliminer le biais qui existait dans la relation entre la présence de cheveux gris et celle d'une fracture

Utilisation des diagrammes causaux

- Définition

Un « diagramme causal » est un « Directed Acyclic Graph » (DAG) :

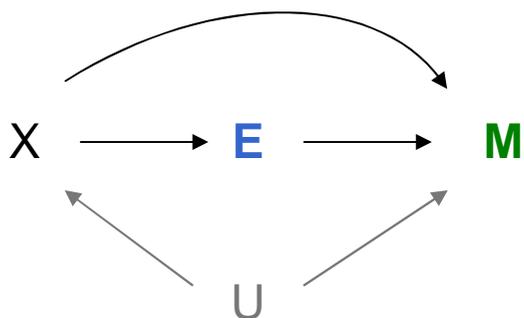
- Graphe qui possède des flèches qui sont dirigées, reliant des expositions et la maladie
- Graphe qui ne doit pas être cyclique (pas de retour à la case départ possible)

- Direction de chaque flèche

- La flèche unidirectionnelle signifie qu'il existe un lien hypothétiquement causal entre les deux variables qu'elle relie
- Pas de flèches bi-directionnelles

Utilisation des diagrammes causaux

- Illustration et terminologie



- M est la maladie, X, E, et U sont trois expositions
- X et E sont mesurées, et U est non mesurée (unknown exposure)
- M est une conséquence de X, E, et U
- U est une cause commune de X et M
- E est sur le chemin causal entre X et M
- X a un lien direct sur M, et un lien indirect passant par E

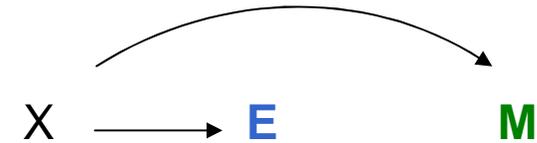
Repérer un biais avec les DAGs

- Méthode
 - Placer l'exposition d'intérêt principal **E** et la maladie **M** sur la feuille de papier
 - Placer toutes les autres expositions qui pourraient éventuellement avoir un lien avec E et/ou M
 - NE PAS METTRE de flèche entre E et M
 - Placer les flèches reliant les expositions entre elles, et/ou avec la maladie
 - Vérifier visuellement (voir méthode plus loin) si la non prise en compte d'une exposition autre que E conduit à une association (fallacieuse) entre E et M

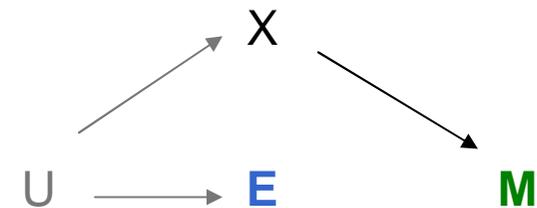
DAGs et biais de confusion (1)

- Situations de biais de confusion

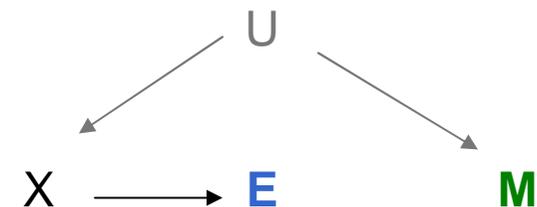
1) E et M ont une cause commune (X)



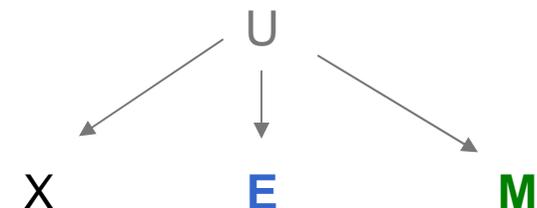
2) E et M ont une cause commune (U), X étant sur le chemin causal entre U et M



3) E et M ont une cause commune (U), X étant sur le chemin causal entre U et E



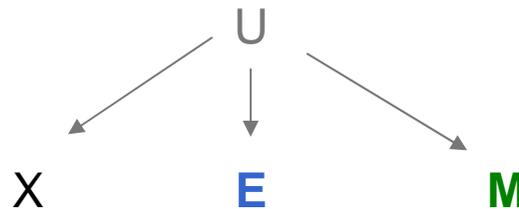
4) E et M ont une cause commune (U), X étant une conséquence de U



DAGs et biais de confusion (1)

- Commentaires

- Dans les 3 dernières situations, la cause de confusion est U, mais X est considérée comme un facteur de confusion
- Dans ces 4 situations, X doit être prise en compte pour diminuer le biais de confusion dû à X (1^{ère} situation) ou dû à U (2^{ème}, 3^{ème}, et 4^{ème} situations)
- Dans la 4^{ème} situation, prendre en compte X n'éliminera pas complètement le biais de confusion dû à U (contrairement aux 3 autres situations), sauf si U et X sont extrêmement liées entre elles...



Le « collider », un vrai danger épidémiologique ! (1)

- Définition

Un collider de 2 expositions est une conséquence commune de ces 2 expositions

(Un collider peut aussi être une conséquence d'une exposition et de la maladie)

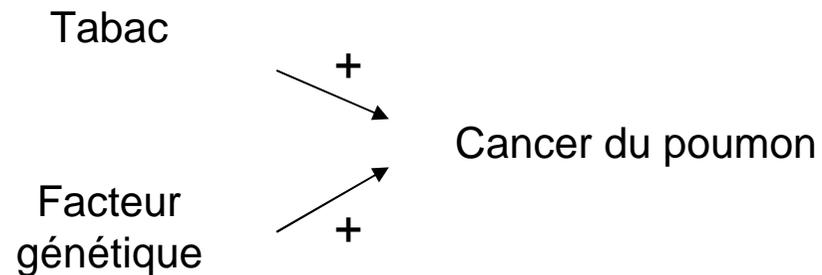


- Pourquoi est-ce un danger ?

Prendre en compte un collider de E_1 et E_2 dans une analyse statistique comme si c'était un facteur de confusion (sélection, stratification, ajustement) va **créer** une association entre E_1 et E_2 !

Le « collider », un vrai danger épidémiologique ! (2)

- Illustration – Présentation



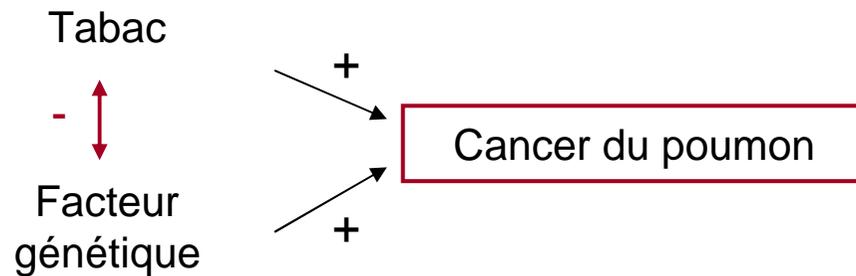
On suppose qu'il n'existe aucun lien entre « tabac » et « facteur génétique »

- Stratifications sur « cancer du poumon »

- Parmi les sujets **sans** cancer, les sujets qui fument ont moins de chance d'avoir le facteur génétique que ceux qui ne fument pas (association négative)
- Parmi les sujets **avec** cancer, les sujets qui ne fument pas ont plus de chance d'avoir le facteur génétique que ceux qui fument (association négative)

Le « collider », un vrai danger épidémiologique ! (3)

- Illustration – Conséquence de prendre en compte un collider



Création d'une association (négative) entre « tabac » et « facteur génétique » quand on stratifie sur une conséquence (collider) de ces deux expositions (le cancer du poumon)

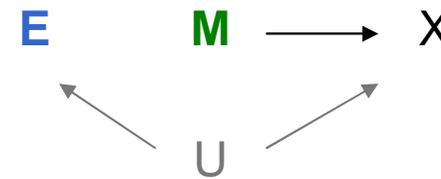
Un collider n'est PAS un facteur de confusion (1)

- Situations simples où l'on ne doit pas prendre en compte X dans les analyses

1) E et M ont une conséquence commune (X)



2) U et M ont une conséquence commune (X),
et U est une cause de E



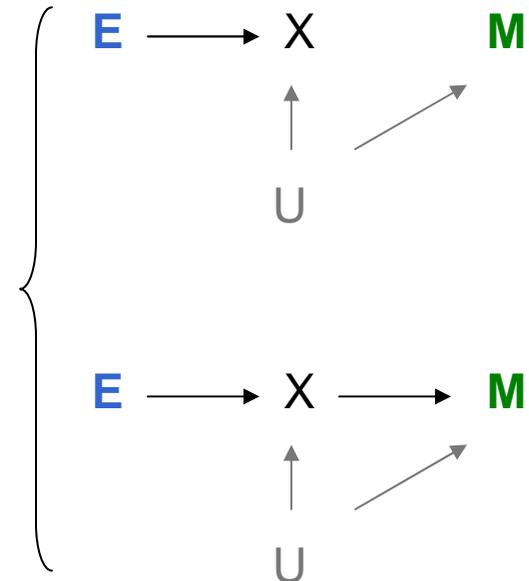
Remarque : situations simples à éviter en interdisant de prendre en compte une conséquence de M

Un collider n'est PAS un facteur de confusion (2)

- Situations plus compliquées où l'on ne doit pas prendre en compte X dans les analyses

X est une conséquence de E, et éventuellement une cause de M (X est alors appelé un facteur intermédiaire dans la relation causale entre E et M)

Cas de figure qui pourrait arriver lorsque l'on veut évaluer l'effet direct de E sur M, effet ne passant pas par X

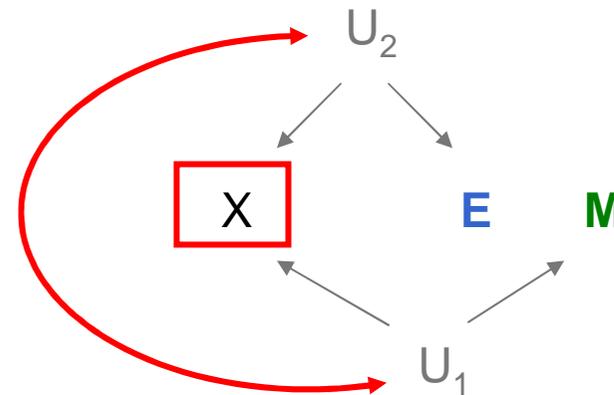


Remarque : prévenir des dangers si prise en compte d'un facteur intermédiaire pour évaluer l'effet direct de E sur M

Un collider n'est PAS un facteur de confusion (3)

- Situation très compliquée où l'on ne doit pas prendre en compte X dans les analyses

X est un indicateur de risque de maladie, et partage une cause commune avec E (U_2)



Remarques :

- Situation difficile à prévoir
- Raison pour laquelle X n'a pas été prise en compte dans les analyses encore plus difficile à justifier dans les parties méthodes des articles !

En conclusion (de l'introduction) sur les DAGs

- Les DAGs (ou diagrammes causaux) permettent...
 - ... de poser le problème (exposition d'intérêt, expositions dont on veut tenir compte, maladie)
 - ... de réfléchir aux hypothèses que l'on fait quant à la temporalité des relations
 - ... de vérifier la pertinence de la prise en compte de variables (facteurs de confusion, conséquences de la maladie, facteurs intermédiaires, collider, ...)
- Les DAGs ne permettent pas ...
 - ... de quantifier la force des associations, réelles ou créées par des prises en compte non recommandées de certaines variables
 - ... de savoir de combien on réduit ou augmente les biais par des prises de compte de variables

Bibliographie

Pearl J. **Causal Diagrams for Empirical Research.** *Biometrika* 1995,82:669-688.

Greenland S, Pearl J, Robins JM. **Causal diagrams for epidemiologic research.** *Epidemiology* 1999,10:37-48.

Hernan MA, Hernandez-Diaz S, Werler MM, Mitchell AA. **Causal knowledge as a prerequisite for confounding evaluation: an application to birth defects epidemiology.** *Am J Epidemiol* 2002,155:176-184.

Hernan MA, Hernandez-Diaz S, Robins JM. **A structural approach to selection bias.** *Epidemiology* 2004,15:615-625.

Hernan MA, Cole SR. **Invited Commentary: Causal Diagrams and Measurement Bias.** *Am J Epidemiol* 2009,170:959-962; discussion 963-954