



Biais et Causalité

Utilisation des

« *Directed Acyclic Graphs* »

Axel Finckh et Anne Lübbecke

Plan

- 1. Quelques rappels sur les notions de facteurs de confusion**
- 2. Les « Directed Acyclic Graphs » (DAGs) (Axel)**
 - Généralités
 - "Les règles et procédures" des DAGs
 - Quelques exemples sur l'utilisation des DAGs pr identifier des confounders
- 3. Quelques rappels sur les notions de biais (Anne)**
- 4. L'utilisation des DAGs pour identifier les bias (Anne)**
 - Bias dans un DAG
 - Quelques exemples sur l'utilisation des DAGs pr identifier des bias

Motivations pour des modèles graphiques

- Nos méthodes statistiques permettent d'étudier des associations (facteurs de risque ou des facteurs pronostiques). Mais **associations \neq cause**. Nous ignorons souvent quelles associations sont causales.
- But des modèles graphiques ➡ rapprocher le modèle conceptuel, mécanistique, du modèle utilisé pour l'analyse des données
 - Modèles graphiques peuvent être utilisés pour représenter les relations causales, connues ou supposées.
- Modèles graphiques ➡ non plus une équation, mais un système d'équations. Ce système d'équations décrit la «*structure*»; c.à.d. les liens de causalité potentiels, d'où leur dénomination « d'équations structurelles »

Quelques rappels sur les notions de biais, de 'causalité' et de 'confounding'

Questions à se poser pour interpréter des résultats épidémiologique

- ? N'y a-t-il pas de biais dans l'étude (i.e. Dr Eric Gerstel...)
- ? Le résultat n'est-il pas du au hasard
- ? Le résultat n'est-il pas dû à autre chose
- ? L'association mesurée est-elle causale

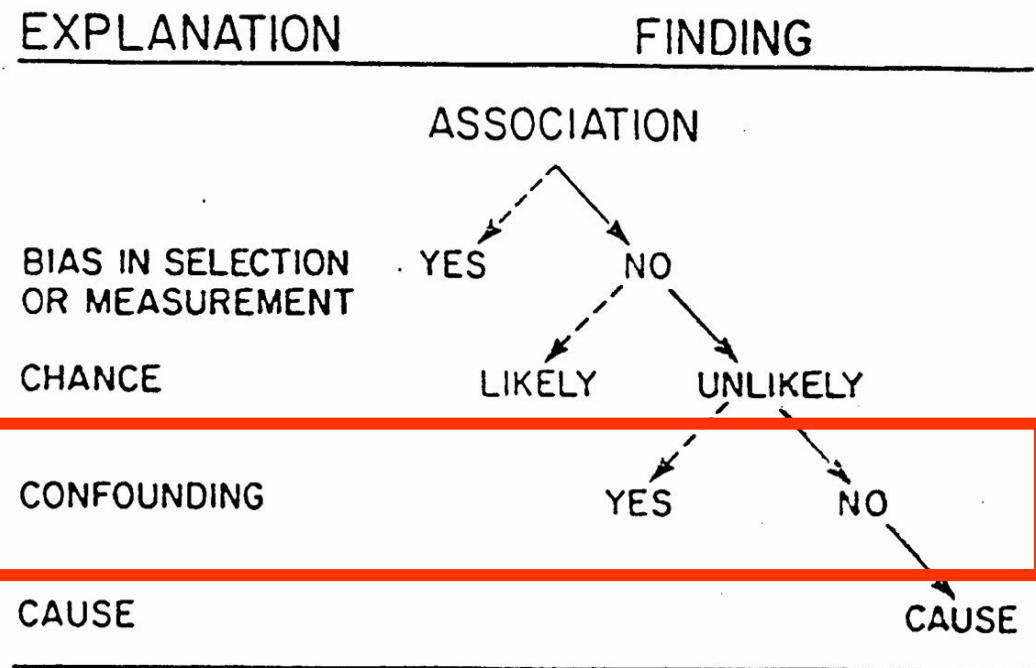
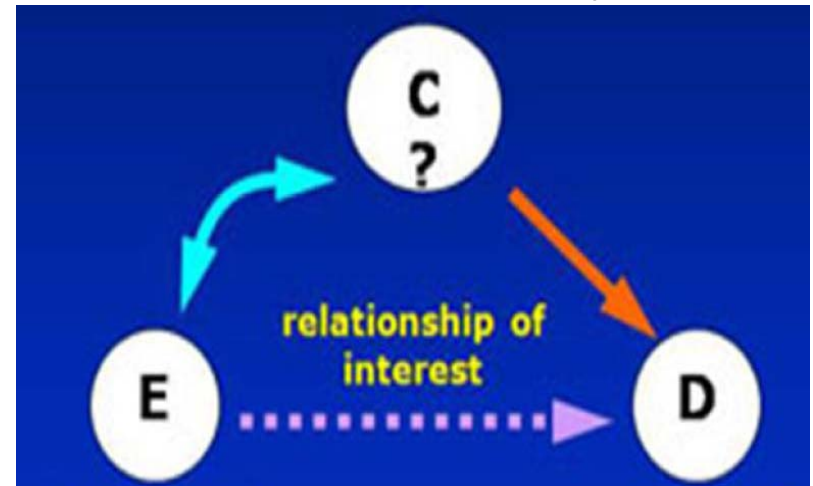


Figure 11.4. Association and cause.

Quelques rappels sur les notions de 'confounding'

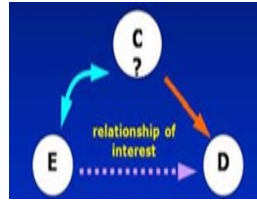
- En épidémiologie: Distingue classiquement 3 types de variables:

- l'exposition (E)
- l'outcome (O)
- les 'confounders' (C)



- Les deux premiers (E, O) sont généralement clairement définis, les facteurs de confusion (C) plus flou ...

Identification des facteurs confondants

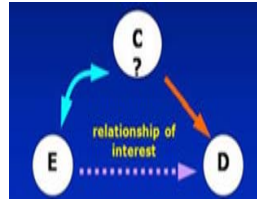


Approches pour identifier des facteurs confondants:

■ Basé sur des **associations statistiques**

1. Procédures 'automatiques' de sélection des variables :
« *Stepwise selection* » (hypothèse sous-jacente: tous les facteurs confondants importants seront sélectionnés...)
2. Comparaison entre les '*crude estimates*' et '*adjusted estimates*': « *10% change rule* » (hypothèse sous-jacente: seuls les facteurs confondants ayant un impact substantiel sur l'estimation valent la peine d'être intégré au modèle ...)

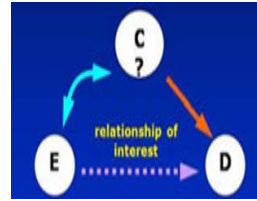
Identification des facteurs confondants



Approches pour identifier des facteurs confondants:

- Basé sur des **critères épidémio. objectifs**
 1. Le facteur C est un facteur de risque connu de la maladie étudiée, indépendamment de l'exposition (E)
 2. Le facteur C est associé à E
 3. Le facteur C n'est pas une conséquence de E
(« *is not on the causal pathway* », « *pas étape intermédiaire dans la chaîne causale* »...)

Identification des facteurs confondants



Toutes ces 3 approches d'identification de facteurs C peuvent induire des biais:

- Par omission des facteurs C importants
- Par ajustements pour facteurs qui ne sont pas des C

➔ un exemple tiré de **AJE** (Miguel Hernàn et al.):



American Journal of Epidemiology
Copyright © 2002 by the Johns Hopkins Bloomberg School of Public Health
All rights reserved

**Causal Knowledge as a Prerequisite for Confounding Evaluation:
An Application to Birth Defects Epidemiology**

Plan

1. Quelques rappels sur les notions de facteurs de confusion
2. **Les « Directed Acyclic Graphs » (DAGs) (Axel)**
 - Généralités
 - "Les règles et procédures" des DAGs
 - Quelques exemples sur l'utilisation des DAGs pr identifier des confounders
3. Quelques rappels sur les notions de biais (Anne)
4. L'utilisation des DAGs pour identifier les bias (Anne)
 - Bias dans un DAG
 - Quelques exemples sur l'utilisation des DAGs pr identifier des bias

« *Directed Acyclic Graphs* » (DAGs)

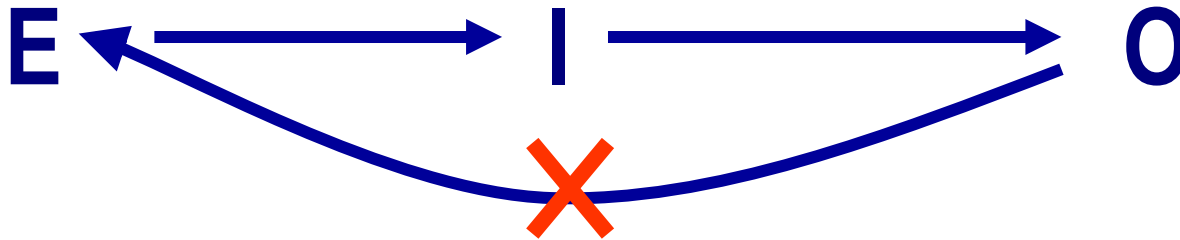
- DAGs link variables (*nodes*) by arrows (*edges*) that represent a direct causal effect
- An arrow placed with its base at E and head at O indicates E causes O. If unsure about causation, still leave an arrow (don't want to assume *a priori* no association)



- **RULE 1:** No arrow indicates no causal relation (Stat: independent conditional on a common cause)
- Because causes precede their effect, DAGs are asymmetric:
 $x \rightarrow y$ is not consistent with $y \rightarrow x$

“Directed Acyclic Graphs” (DAGs)

- A directed acyclic graph is a directed graph that contains **no cyclic paths**
- Temporal ordering: all arrows go from left → right



- The path $E \rightarrow I \rightarrow O \rightarrow E$ is labeled “cyclic” as we move from E to O, but then return to O by way of E.
- Causation doesn’t go backwards → No backward arrows

“Directed Acyclic Graphs” (DAGs)

- DAGs can be causal or statistical graphs
- All causal graphs are also statistical, but not vice versa
- **Statistical Graphs** are defined by Statistical Markov Assumption: Any variable that is not a descendant of (arrow) a given variable E will be independent of E (conditional on E 's parents)
- **Causal Graphs**: A representation that includes all causal relations (measured and unmeasured variables !)



Association = Causation, when all other causal paths are blocked

“Directed Acyclic Graphs” (DAGs)

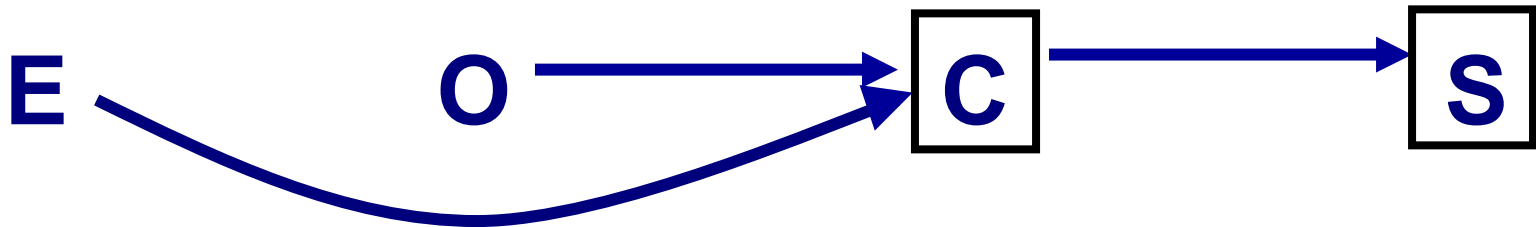
- **Definition of a Path:** an arrow-based route between variables in a DAG (associations can go both ‘*backward*’ and ‘*forward*’; but causation has a temporal order)



A path is blocked :

- If no arrow
- if 2 arrowheads on the path ‘collide’ at some variable along the path (“*collider*”). I.e.: $E \rightarrow O \leftarrow U$
- if a variable (“*non-collider*”) has been conditioned on. (i.e.. stratified) I.e.: $U \rightarrow \boxed{E} \rightarrow O$

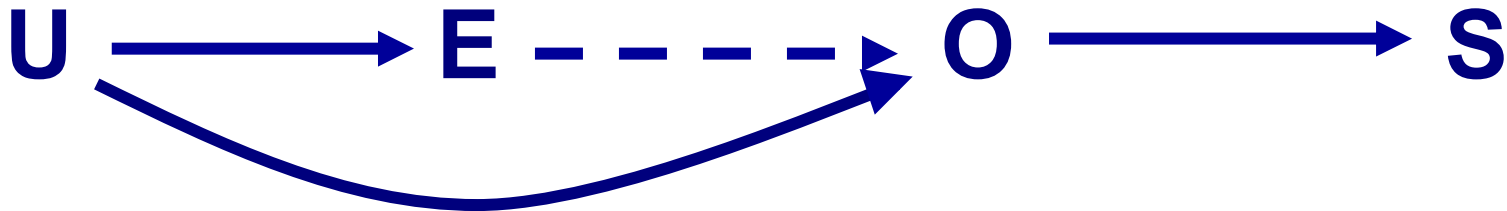
“Directed Acyclic Graphs” (DAGs)



- A **path is not blocked** (between E and O):
 - if conditions on a “*collider*”. I.e.: $E \rightarrow \boxed{C} \leftarrow O$
 - if conditions on “*descendant*” of a “*collider*”.
I.e.: path between U and E is open !

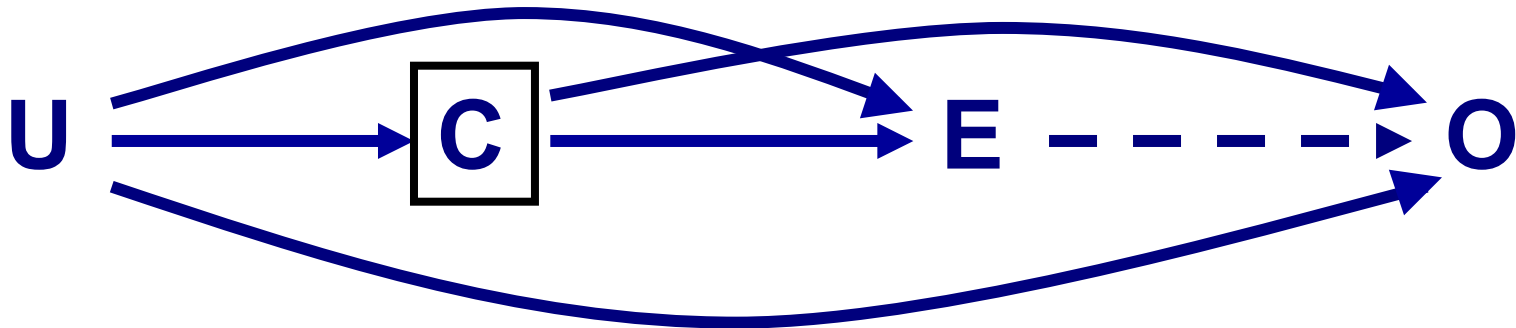
“Directed Acyclic Graphs” (DAGs)

- **‘D-separation’**: Graphical rules used to determine whether 2 variables are independent of each other



- **Consequences**: In general ...
 - Exposures (causes, parents) and outcomes (effects, children) are not independent and vice versa
 - 2 variables are associated if they share a common cause
 - A common effect, does not imply that 2 causes are associated
 - 2 causes become associated if we stratify on a common effect

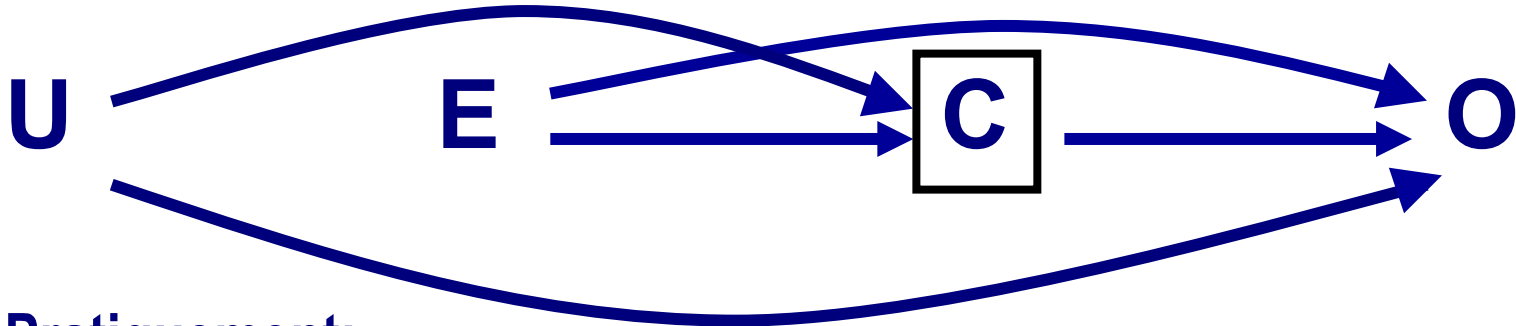
Utiliser les DAGs pour examiner d'év. facteurs C



■ Pratiquement:

- Dessiner un DAG avec toutes les causes imaginables pour un Outcome (O) donné
- Enlever les flèches entre les variables lorsque nous *savons* (!) qu'il n'y a pas d'association
- Examiner si il existe un autre chemin (*'backdoor path'*) entre notre exposition E et notre outcome O, autrement dit voir si l'association entre $E \rightarrow O$ peut être expliqué par un autre chemin, c'ad un facteur confondant
- Ici, ajuster pour C permettra de contrôler pour facteur de confusion

DAGs & “variables intermédiaires”



■ Pratiquement:

- Dessiner un DAG avec toutes les causes imaginables pour un Outcome (O) donné et enlever les flèches entre les variables lorsque nous *savons* (!) qu'il n'y a pas d'association
- Examiner si il existe un autre chemin ('backdoor path') entre notre exposition E et notre outcome O, autrement dit voir si l'association entre $E \rightarrow O$ peut être expliqué par un autre chemin, càd un facteur confondant
- Si on avait '*conditionné*' pour C (p.ex. bonne tolérance du médicament), on aurait ouvert un 'backdoor path'; E et O seront associé (« *conditionally associated with strata of C* »)

DAGs pr identifier des facteurs confondants



American Journal of Epidemiology

Copyright © 2002 by the Johns Hopkins Bloomberg School of Public Health

All rights reserved

Causal Knowledge as a Prerequisite for Confounding Evaluation: An Application to Birth Defects Epidemiology

Miguel Hernàn et al.

➔ **Les 3 approches d'identification de facteurs C traditionnels peuvent induire des biais:**

- Par omission des facteurs C importants
- Par ajustements pour facteurs qui ne sont pas des C

Exemple

Periconceptional Folic Acid Exposure and Risk of Occurrent Neural Tube Defects

Martha M. Werler, ScD; Samuel Shapiro, MB, FRCPE; Allen A. Mitchell, MD

- **D:** Mothers of infants with neural tube defect
- **E:** Suppl. of folic acid < the first 2 months
- **C:** Facteur confondant, binaire, pas intermédiaire

Example: Periconceptional Folic Acid Exposure and Risk of Neural Tube Defects

Results - Crude

D = 1 **D = 0**

E = 1

43	239
194	704

Crude OR: 0.65 (0.45 – 0.94)

E = 0

- **D:** Mothers of infants with neural tube defect
- **E:** Suppl. of folic acid < the first 2 months

Example: Periconceptional Folic Acid Exposure and Risk of Neural Tube Defects

Results – Adjusted (C: Confounder Variable)

C=1

D = 1

D = 0

C=0

D = 1

D = 0

E = 1

19	8
100	46

E = 1

24	231
94	658

E = 0

E = 0

- Absence of significant heterogeneity ($p=0.43$) => Pool
- **Adjusted OR: 0.80 (0.53 – 1.20)**

Exemple: Periconceptional Folic Acid Exposure and Risk of NeuralTube Defects

Quel résultat est le le plus adéquat?

1. Sélection 'automatique' de variables:

C significativement associé à D ($p < 0.1$) => intégré au modèle

2. Changement de l'estimateur > 10%:

OR passe de 0.65 à 0.80, càd \uparrow de 23% => va être pris

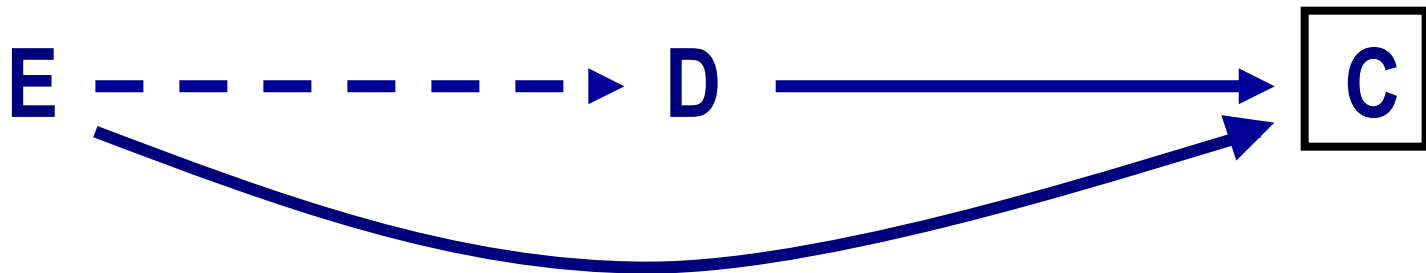
3. Définition standard pour facteurs confondants:

- C est associé avec E: $OR_{CE|D} 0.50 (0.23 - 1.07)$
- C est associé avec D c/o non exposés: $OR_{CD|E=0} 15.2 (10.1 - 23)$
- C n'est pas une variable intermédiaire (not on the causal pathway)

=> va être pris

Exemple: Periconceptional Folic Acid Exposure and Risk of Neural Tube Defects

C: Mort in utero ou Avortement thérapeutique

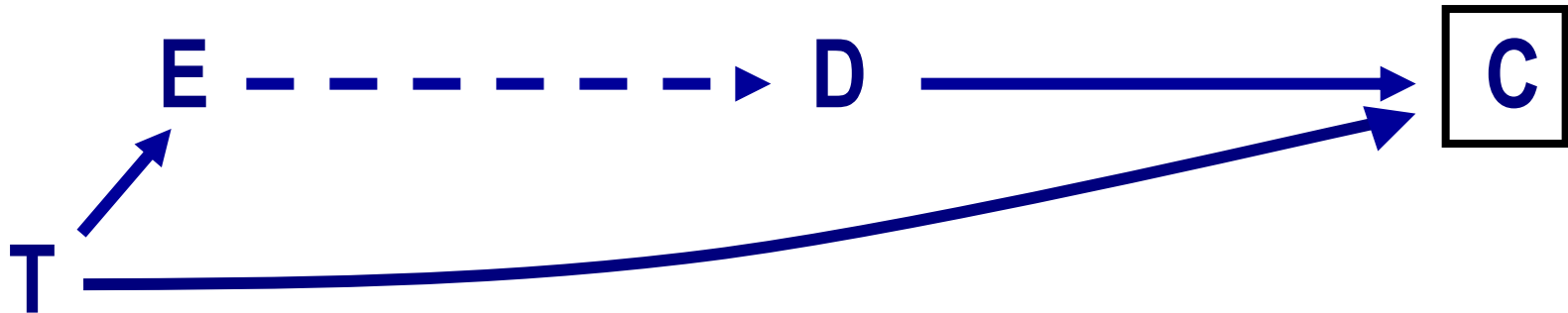


- Carences en ac. folique augmentent le risque de mort in utero
- Les Anomalies du tube neural sont associés avec une ↑ avortements thérapeutiques et mortalité foetale

Exemple: Periconceptional Folic Acid Exposure and Risk of NeuralTube Defects

C: Mort in utero ou Avortement thérapeutique

T: Grosses adolescentes (« Teen »)



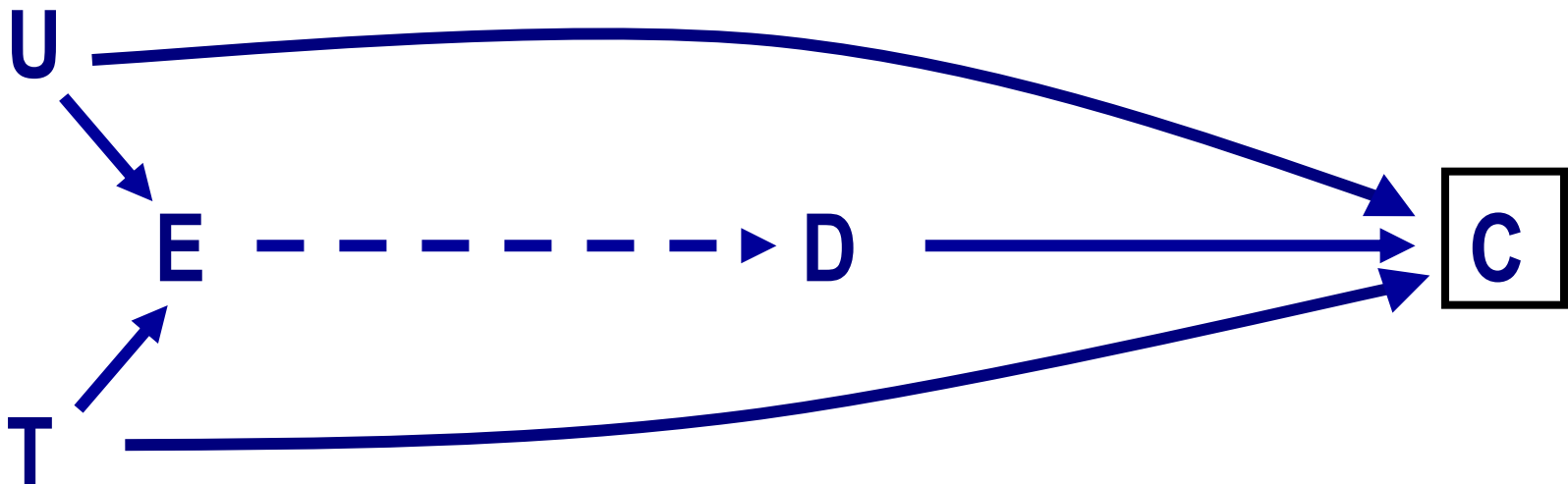
- Grossesses chez adolescentes ↑ avortements & naissances difficiles ...
- Grossesses chez adolescentes associés avec ↓ de suppléments vitaminiques

Exemple: Periconceptional Folic Acid Exposure and Risk of NeuralTube Defects

C: Prise de poids maternel durant la grossesse

T: Grossesse d'adolescentes

U: Facteur génétique: \uparrow prise poids, \downarrow ac. folique



Plan

1. Quelques rappels sur les notions de facteurs de confusion
2. Les « Directed Acyclic Graphs » (DAGs) (Axel)
 - Généralités
 - "Les règles et procédures" des DAGs
 - Quelques exemples sur l'utilisation des DAGs pour identifier des confounders
3. **Quelques rappels sur les notions de biais (Anne)**
4. L'utilisation des DAGs pour identifier les biais (Anne)
 - Biais dans un DAG
 - Quelques exemples sur l'utilisation des DAGs pour identifier des biais

Erreurs systématiques

■ Facteurs de confusion

■ Biais d'information

- « Mis-classification »
- Biais de ré-mémorisation

■ Biais de sélection

Rothman KJ, Greenland S, Lash TL. *Modern Epidemiology*, 3rd edition. 2008

Erreurs systématiques

- Facteurs de confusion

- Biais d'information

- « Mis-classification »

- Biais de ré-mémorisation

- Biais de sélection

Rothman KJ, Greenland S, Lash TL. *Modern Epidemiology*, 3rd edition. 2008

Biais de sélection

■ Biais de sélection – définition:

□ Association **E** → **O** différente entre sujets éligibles (population d'origine) et sujets inclus dans l'analyse *

■ « Berkson's bias »

■ Biais par sélection intentionnelle (« Differential loss-to follow-up, Survivor bias, Self-selection bias »)

■ Sélection inappropriée de témoins (études cas-témoins)

*Rothman KJ, Greenland S. *Modern Epidemiology* 2nd edition. 1998

*Szklo MO, Nieto FJ. *Epidemiology. Beyond the Basics*. 2000

Plan

1. Quelques rappels sur les notions de facteurs de confusion
2. Les « Directed Acyclic Graphs » (DAGs) (Axel)
 - Généralités
 - "Les règles et procédures" des DAGs
 - Quelques exemples sur l'utilisation des DAGs pour identifier des confounders
3. Quelques rappels sur les notions de biais (Anne)
4. L'utilisation des DAGs pour identifier les biais (Anne)
 - Biais dans un DAG
 - Quelques exemples sur l'utilisation des DAGs pour identifier des biais

Directed acyclic graphs (DAGs)

■ DAGs:

- Très utile pour représenter des structures causales
- Association peut être:

- **Cause et effet:**



Association causale

Hernan M et al. A structural approach to selection bias. *Epidemiology*, 2004
Glymour M, Greenland S, Causal diagrams, *Modern Epidemiology*, 3rd edition 2008

Directed acyclic graphs (DAGs)

- Cause et effet:



« *Reverse causation* »

Directed acyclic graphs (DAGs)

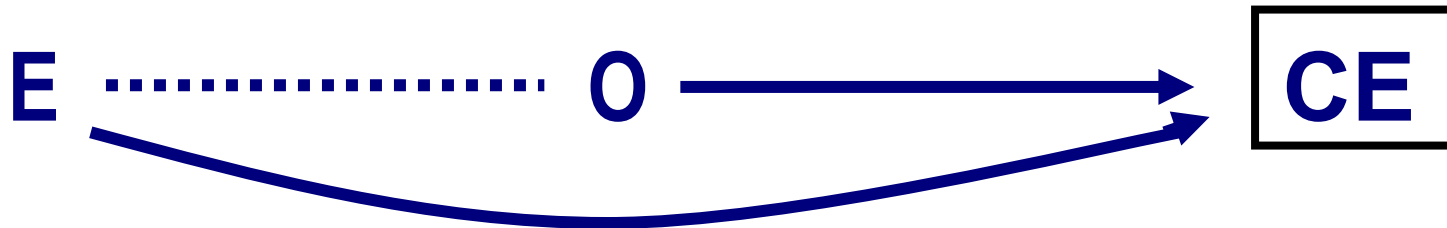
- **Causes communes:** Si **E** et **O** ont une cause commune (**C**), ceci peut résulter en une association entre les deux qui n'est pas causale



« *Confounding* »

Directed acyclic graphs (DAGs)

- **Effets communs:** **E** et **O** qui ont un effet commun (**CE**), seront associé si l'effet est estimée dans des sous-groupes de **CE** («conditioning on **CE** »)



Biais de sélection, « Collider bias »

Biais de sélection

- Berkson's bias (Collider bias):

- Maladies A et B non-associées dans la population, mais associées parmi patients hospitalisés si les deux maladies influencent la probabilité d'admission à l'hôpital

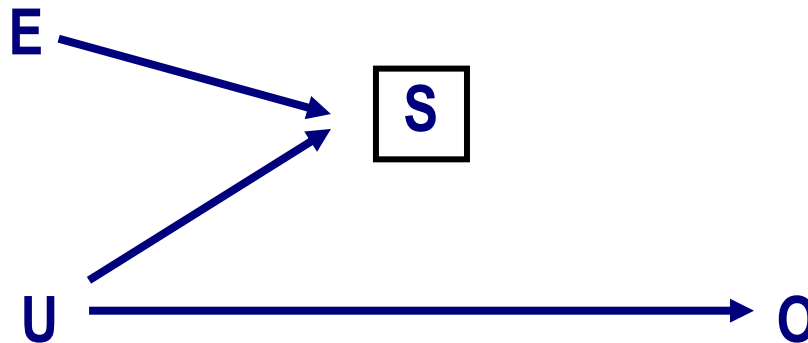


- Autre exemple: Estrogen Endometrial cancer → Bleeding
-
- A smaller causal diagram showing 'Estrogen' connected to 'Endometrial cancer' by a dotted line, and 'Endometrial cancer' connected to 'Bleeding' by a solid arrow. A curved solid arrow also points from 'Estrogen' to 'Bleeding', representing a collider effect.

Biais de sélection

- Biais par sélection intentionnelle

- « Differential loss to follow-up »
- « Survivor bias »
- « Self-selection bias »



- Conceptualisation de la sélection par une variable **S** = Sélection oui / non

Biais de sélection

- Sélection inappropriée de témoins (études cas-témoins)



E = Postmenopausal estrogen

O = Risk of myocardial infarction

F = Women with hip fracture

S = Selection for case-control study from the population (yes/no)

Témoins prenaient de l'estrogène moins souvent!

Limitations

- Pas d'estimation de l'ampleur ni de la direction du biais
- Pas de quantification de l'ampleur de la relation examinée
- Décision quelle variable inclure est parfois difficile

Conclusions

- DAGs offrent un schéma des structures causales présumées
- Model visuel flexible
- Le défi d'utiliser des DAGs peut être bénéfique dans la mesure que ça nous force à clarifier nos hypothèses
- Le vraie diagramme causal est inconnu, mais faire un DAG nous permet d'identifier les sources de biais et d'incertitudes dans nos observations et conclusions