

Using DAGs and conditional independence to understand complex models in Epidemiology

S.Geneletti

Imperial College

07/02/2008

- 1 How to use directed acyclic graphs to formulate questions and understand complex systems in Epi

- 1 How to use directed acyclic graphs to formulate questions and understand complex systems in Epi
- 2 How to use them as formal tools to describe a problem

- 1 How to use directed acyclic graphs to formulate questions and understand complex systems in Epi
- 2 How to use them as formal tools to describe a problem
- 3 How to make causal inference where possible

- 1 Definition of DAGs and conditional independence

Outline

- 1 Definition of DAGs and conditional independence
- 2 Examples of how to use DAGs to determine whether something is a confounder or an intermediate

- 1 Definition of DAGs and conditional independence
- 2 Examples of how to use DAGs to determine whether something is a confounder or an intermediate
- 3 Introduce the Decision-Theoretic framework for causal inference

- 1 Definition of DAGs and conditional independence
- 2 Examples of how to use DAGs to determine whether something is a confounder or an intermediate
- 3 Introduce the Decision-Theoretic framework for causal inference
- 4 Examples of how to use DAGs to understand gene-environment interactions
- 5 Conclusions

Definition

- Directed Acyclic Graphs

Definition

- Directed Acyclic Graphs
- Nodes represent variables

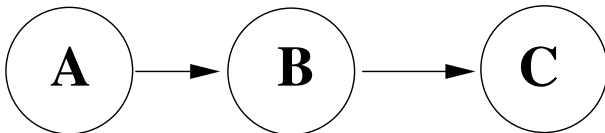
Definition

- Directed Acyclic Graphs
- Nodes represent variables
- edges all directed

Definition

- Directed Acyclic Graphs
- Nodes represent variables
- edges all directed
- no loops $A \rightarrow B \rightarrow A$

Example:



Conditional Independence

$$p(A, C|B) = p(A|B) \times p(C|B) \Rightarrow A \perp\!\!\!\perp C|B^{[1]}$$

Conditional Independence

$$p(A, C|B) = p(A|B) \times p(C|B) \Rightarrow A \perp\!\!\!\perp C|B^{[1]}$$

DAGs

- Encode conditional independence (CI) - probability statement
- lack of an edge means no dependence

Conditional Independence

$$p(A, C|B) = p(A|B) \times p(C|B) \Rightarrow A \perp\!\!\!\perp C|B^{[1]}$$

DAGs

- Encode conditional independence (CI) - probability statement
- lack of an edge means no dependence
- an **arrow is not intrinsically causal**

Conditional Independence

$$p(A, C|B) = p(A|B) \times p(C|B) \Rightarrow A \perp\!\!\!\perp C|B^{[1]}$$

DAGs

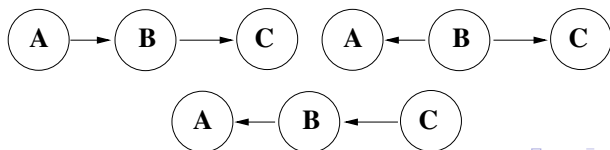
- Encode conditional independence (CI) - probability statement
- lack of an edge means no dependence
- an **arrow is not intrinsically causal**
- Can read CI's from DAGs
- More than one can encode same CI - $A \perp\!\!\!\perp C|B$

Conditional Independence

$$p(A, C|B) = p(A|B) \times p(C|B) \Rightarrow A \perp\!\!\!\perp C | B^{[1]}$$

DAGs

- Encode conditional independence (CI) - probability statement
- lack of an edge means no dependence
- an **arrow is not intrinsically causal**
- Can read CI's from DAGs
- More than one can encode same CI - $A \perp\!\!\!\perp C | B$



Story

What is the role of **C-reactive protein (CRP)** w.r.t **cardio-vascular (CV) events**, **low-density lipoprotein cholesterol (LDL)** and another variable **X** representing unknown/suspected factors

Story

What is the role of **C-reactive protein (CRP)** w.r.t **cardio-vascular (CV)** events, **low-density lipoprotein cholesterol (LDL)** and another variable X representing unknown/suspected factors

- 1 Is CRP an **intermediate** on path from LDL to CV?

Story

What is the role of **C-reactive protein (CRP)** w.r.t **cardio-vascular (CV)** events, **low-density lipoprotein cholesterol (LDL)** and another variable **X** representing unknown/suspected factors

- 1 Is CRP an **intermediate** on path from LDL to CV?
- 2 Is it simply **associated** to LDL via some other variable **X**?

Story

What is the role of **C-reactive protein (CRP)** w.r.t **cardio-vascular (CV)** events, **low-density lipoprotein cholesterol (LDL)** and another variable X representing unknown/suspected factors

- 1 Is CRP an **intermediate** on path from LDL to CV?
- 2 Is it simply **associated** to LDL via some other variable X?
- 3 Is it on an **alternative pathway** from X to CV?

Story

What is the role of **C-reactive protein (CRP)** w.r.t **cardio-vascular (CV)** events, **low-density lipoprotein cholesterol (LDL)** and another variable X representing unknown/suspected factors

- 1 Is CRP an **intermediate** on path from LDL to CV?
- 2 Is it simply **associated** to LDL via some other variable X?
- 3 Is it on an **alternative pathway** from X to CV?
- 4 Is it a **confounder** of the relationship between CV and X??

Story

What is the role of **C-reactive protein (CRP)** w.r.t **cardio-vascular (CV)** events, **low-density lipoprotein cholesterol (LDL)** and another variable X representing unknown/suspected factors

- 1 Is CRP an **intermediate** on path from LDL to CV?
- 2 Is it simply **associated** to LDL via some other variable X?
- 3 Is it on an **alternative pathway** from X to CV?
- 4 Is it a **confounder** of the relationship between CV and X??

Question: How can we distinguish between the above scenarios?

Idea

- 1 See what conditional independences could describe the problem and see if they hold in data
- 2 this can tell us about the role of CRP.

Idea

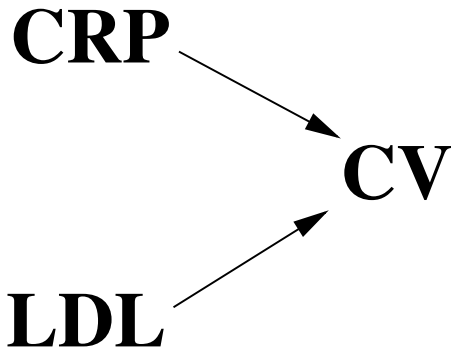
- 1 See what conditional independences could describe the problem and see if they hold in data
- 2 this can tell us about the role of CRP.
- 3 Assume: CV “downstream” as it is final event
- 4 Assume LDL known cause of CV so that $LDL \rightarrow CV$
- 5 These are temporal/causal assumption to reduce number of possibilities
- 6 Focus on role of CRP

Scenario 1:

- 1 Both LDL and CRP assoc to risk of CV

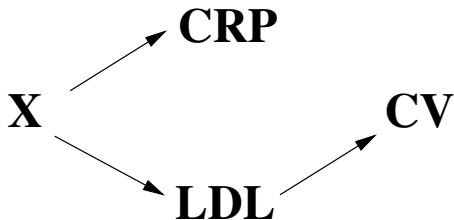
Scenario 1:

- 1 Both LDL and CRP assoc to risk of CV



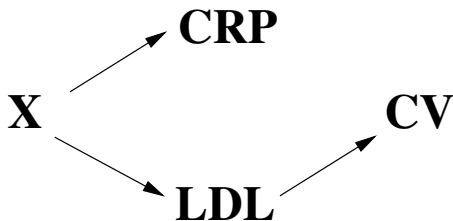
Question

- 1 Is this telling the whole story?
- 2 Perhaps they are associated via X i.e.
- 3 **Scenario 2:**



Question

- 1 Is this telling the whole story?
- 2 Perhaps they are associated via X i.e.
- 3 **Scenario 2:**



Conditional independences:

(i) $CRP \perp\!\!\!\perp LDL \mid X$

(ii) $CV \perp\!\!\!\perp CRP \mid (LDL, X)$? CV does not depend “directly” on CRP

Scenario 3:

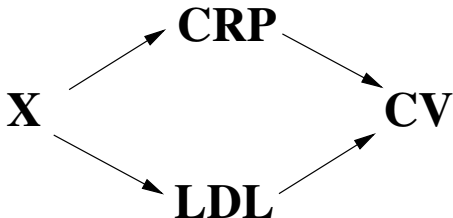
- 1 Maybe CRP is on an alternative path from X to CV?

Scenario 3:

- 1 Maybe CRP is on an alternative path from X to CV?
(i) $CRP \perp\!\!\!\perp LDL | X$
- 2 but this time CV depends on CRP

Scenario 3:

- 1 Maybe CRP is on an alternative path from X to CV?
(i) $CRP \perp\!\!\!\perp LDL | X$
- 2 but this time CV depends on CRP



Could this be wrong though?

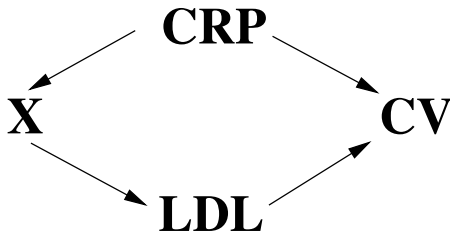
Scenario 4: confounding

- 1 CRP could be a **confounder** for the assoc between X, LDL and CV
- 2 This could still mean that $CRP \perp\!\!\!\perp LDL | X$ is true

Could this be wrong though?

Scenario 4: confounding

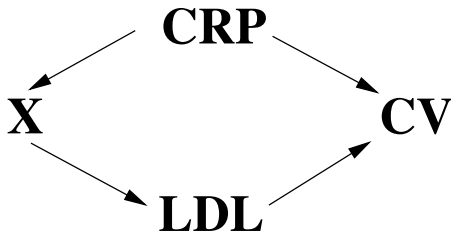
- 1 CRP could be a **confounder** for the assoc between X, LDL and CV
- 2 This could still mean that $CRP \perp\!\!\!\perp LDL | X$ is true



Could this be wrong though?

Scenario 4: confounding

- 1 CRP could be a **confounder** for the assoc between X, LDL and CV
- 2 This could still mean that $CRP \perp\!\!\!\perp LDL | X$ is true



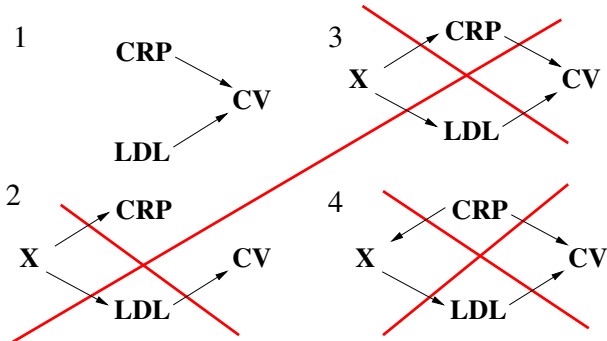
- 3 But now, changing CRP affects LDL via X

Examine the evidence

- 1 From paper [2] CRP and LDL have low ($r=0.13$) but significant ($p=0.005$) correlation

Examine the evidence

- 1 From paper [2] CRP and LDL have low ($r=0.13$) but significant ($p=0.005$) correlation
- 2 We could assume **no association** i.e. $CRP \perp\!\!\!\perp LDL$

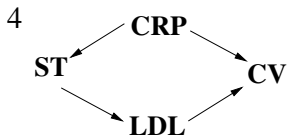
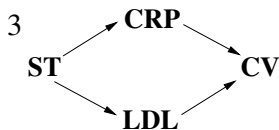
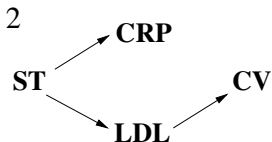
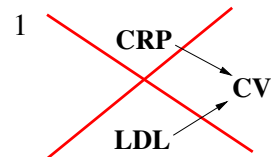


Examine the evidence

- 1 From paper [3] CRP and LDL have low ($r=0.13$) but significant ($p=0.005$) correlation

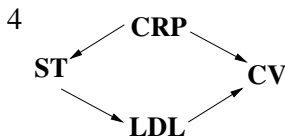
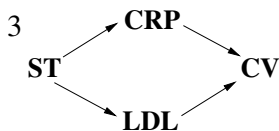
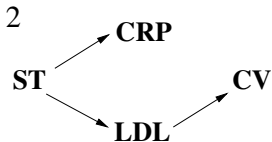
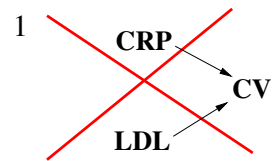
Examine the evidence

- 1 From paper [3] CRP and LDL have low ($r=0.13$) but significant ($p=0.005$) correlation
- 2 Or we could assume there **is an association**



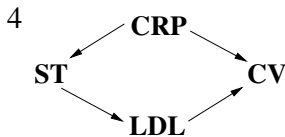
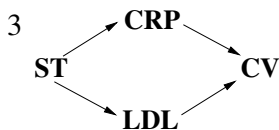
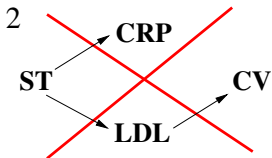
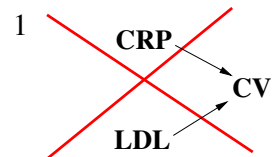
Examine the evidence on Statin trials

- 1 Look at Statin trials^[2,3] for lowering LDL - also lower CRP
- 2 **Statin could be a possible X** as [2] says that CRP and LDL not associated in levels of statin i.e.
 $CRP \perp\!\!\!\perp LDL | ST$ where ST is Statin
- 3 But which scenario reflects the situation?



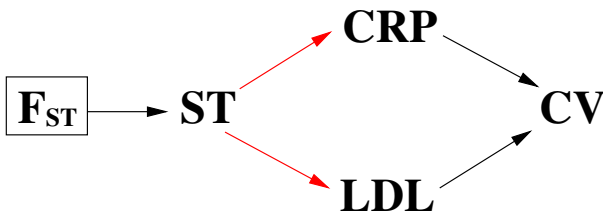
Examine the evidence on Statin trials

- 1 Within levels of LDL, CRP has an impact on CV i.e. $CRP \not\perp CV | (LDL, ST)$
- 2 This means CRP has **direct** association with CV
- 3 Does not mean CRP is not associated to LDL via ST - could it still be a confounder?



Causality

- 1 How do we distinguish between Scenarios 3 & 4?
- 2 We invoke the fact that Statin lowers levels of both CRP and LDL in a **randomised**^[4] trial
- 3 This tells us that $ST \rightarrow CRP$ is causal (for trial popⁿ).



Causality - Formal

- 1 So far looked at DAGs as tools to describe observed associations
- 2 Arrows in DAGs don't mean causality **unless we put it there** - many ways of doing this

Causality - Formal

- 1 So far looked at DAGs as tools to describe observed associations
- 2 Arrows in DAGs don't mean causality **unless we put it there** - many ways of doing this
- 3 Next describe my way of making causal inference using DAGs and CI

Causality - Formal

- 1 So far looked at DAGs as tools to describe observed associations
- 2 Arrows in DAGs don't mean causality **unless we put it there** - many ways of doing this
- 3 Next describe my way of making causal inference using DAGs and CI
- 4 Not the only way - see Jamie Robins, Judea Pearl, Donald Rubin for a selection of approaches

Causality - Formal

- 1 So far looked at DAGs as tools to describe observed associations
- 2 Arrows in DAGs don't mean causality **unless we put it there** - many ways of doing this
- 3 Next describe my way of making causal inference using DAGs and CI
- 4 Not the only way - see Jamie Robins, Judea Pearl, Donald Rubin for a selection of approaches
- 5 I prefer mine because - **no counterfactuals, no determinism, no arrows=causality**

Interventions

- 1 Basic idea - **cause = intervention**
- 2 i.e. without interventions cannot have causality!!!

Interventions

- 1 Basic idea - **cause = intervention**
- 2 i.e. without interventions cannot have causality!!!
- 3 What if you have only observational data?

Interventions

- 1 Basic idea - **cause = intervention**
- 2 i.e. without interventions cannot have causality!!!
- 3 What if you have only observational data?
- 4 **Loophole:** if you can show that in a particular observational setting the behaviour of the system is the same as under intervention then we can talk about causality
- 5 can do this using CI's

Interventions

- 1 Basic idea - **cause = intervention**
- 2 i.e. without interventions cannot have causality!!!
- 3 What if you have only observational data?
- 4 **Loophole:** if you can show that in a particular observational setting the behaviour of the system is the same as under intervention then we can talk about causality
- 5 can do this using CI's
- 6 Framework called **Decision theoretic** (based on^[5])

Interventions

- 1 Basic idea - **cause = intervention**
- 2 i.e. without interventions cannot have causality!!!
- 3 What if you have only observational data?
- 4 **Loophole:** if you can show that in a particular observational setting the behaviour of the system is the same as under intervention then we can talk about causality
- 5 can do this using CI's
- 6 Framework called **Decision theoretic** (based on^[5])

Motivation

- 1 **no cause in no cause out**^[6]

Intervention variable

F_X for r.v. X is such that

- 1 $p(X = x | F_X = x, \mathbf{Z}) = 1$ means **set** $X = x$

Intervention variable

F_X for r.v. X is such that

- 1 $p(X = x | F_X = x, \mathbf{Z}) = 1$ means **set** $X = x$
- 2 $F_X = x$ means **intervene upon**
- 3 E.g. drug trials, animal experiments - even gene knockout!

Intervention variable

F_X for r.v. X is such that

- 1 $p(X = x | F_X = x, \mathbf{Z}) = 1$ means **set** $X = x$
- 2 $F_X = x$ means **intervene upon**
- 3 E.g. drug trials, animal experiments - even gene knockout!
- 4 $p(X = x | F_X = \emptyset, \mathbf{Z}) = p(X = x | \mathbf{Z})$ **natural** distribution of X conditional on \mathbf{Z} - set of other variables in problem.

Intervention variable

F_X for r.v. X is such that

- 1 $p(X = x | F_X = x, \mathbf{Z}) = 1$ means **set** $X = x$
- 2 $F_X = x$ means **intervene upon**
- 3 E.g. drug trials, animal experiments - even gene knockout!
- 4 $p(X = x | F_X = \emptyset, \mathbf{Z}) = p(X = x | \mathbf{Z})$ **natural** distribution of X conditional on \mathbf{Z} - set of other variables in problem.
- 5 $F_X = \emptyset$ means **leave alone**
- 6 E.g. observational studies

The problem with confounding

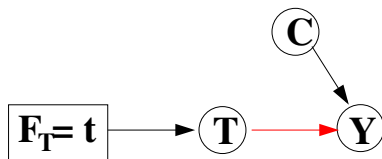
- 1 We want to be able to say biological exposure X causes disease Y

The problem with confounding

- 1 We want to be able to say biological exposure X causes disease Y
- 2 Can do it in labs and petri dishes
- 3 But those are **controlled and artificial situations**

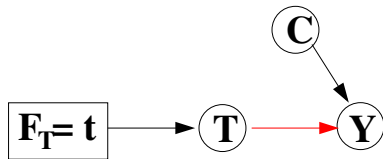
The problem with confounding

- 1 We want to be able to say biological exposure X causes disease Y
- 2 Can do it in labs and petri dishes
- 3 But those are **controlled and artificial situations**
- 4 How do we know red arrow is causal - for sure from randomisation

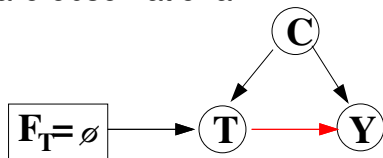


The problem with confounding

- 1 We want to be able to say biological exposure X causes disease Y
- 2 Can do it in labs and petri dishes
- 3 But those are **controlled and artificial situations**
- 4 How do we know red arrow is causal - for sure from randomisation



- 5 but what if data are observational?

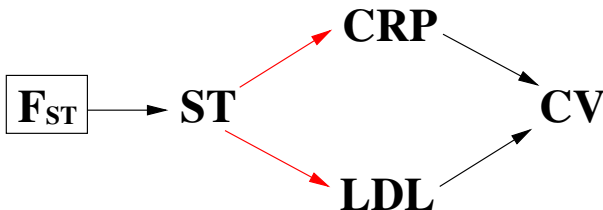


The problem with confounding continued

- 1 Similar issue with genes and saying gene G causes disease Y
- 2 like saying “being a woman causes breast cancer” ?
- 3 Our bodies are so complex, our understanding so tenuous
- 4 Caution needs to be taken when saying associations are causal
- 5 Not ruling it out! but **BE CAREFUL!** especially when inclined to interpret arrows in DAGs as causal.

Intervention

- 1 There are many trials of Statin treatments for CV patients^[2,3,4]
- 2 Results from trials mean that at least red arrows are potentially causal
- 3 Bearing in mind that we are looking at a limited population of people already suffering from CV - no guarantee they are generalisable to healthy people.



Instruments

When intervention is not possible and only have observational studies:

- 1 Using **mendelian randomization**^[7] as an **instrumental variable** is gaining popularity in epi^[8]

Instruments

When intervention is not possible and only have observational studies:

- 1 Using **mendelian randomization**^[7] as an **instrumental variable** is gaining popularity in epi^[8]
- 2 **Idea:** Find a variable A related to CRP, that is not assoc to LDL or CV directly.

Instruments

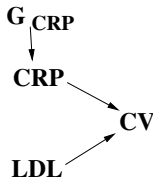
When intervention is not possible and only have observational studies:

- 1 Using **mendelian randomization**^[7] as an **instrumental variable** is gaining popularity in epi^[8]
- 2 **Idea:** Find a variable A related to CRP , that is not assoc to LDL or CV directly.
- 3 E.g. some people naturally do not produce as much CRP - the gene (G_{CRP}) that controls production is faulty

Instruments

When intervention is not possible and only have observational studies:

- 1 Using **mendelian randomization**^[7] as an **instrumental variable** is gaining popularity in epi^[8]
- 2 **Idea:** Find a variable A related to CRP , that is not assoc to LDL or CV directly.
- 3 E.g. some people naturally do not produce as much CRP - the gene (G_{CRP}) that controls production is faulty
- 4 Provided G_{CRP} is not assoc to LDL or CV via other genes or confounders,
- 5 then we can **use G_{CRP} as a substitute for intervention**



GEI example, Parkinson's, exposure to pesticides and the DJ-1 gene

Based on work with P.Vineis and V.Gallo online at www1.imperial.ac.uk/medicine/about/divisions/ephpc/eph/eph_publications/techreports/

Story

- 1 Some evidence of association between **Parkinson's** and exposure to pesticides in particular **Paraquat**^[9]

GEI example, Parkinson's, exposure to pesticides and the DJ-1 gene

Based on work with P.Vineis and V.Gallo online at www1.imperial.ac.uk/medicine/about/divisions/ephpc/eph/eph_publications/techreports/

Story

- 1 Some evidence of association between **Parkinson's** and exposure to pesticides in particular **Paraquat**^[9]
- 2 Also some evidence of association between **Parkinson's** and **DJ-1 gene**^[10]

GEI example, Parkinson's, exposure to pesticides and the DJ-1 gene

Based on work with P.Vineis and V.Gallo online at www1.imperial.ac.uk/medicine/about/divisions/ephpc/eph/eph_publications/techreports/

Story

- 1 Some evidence of association between **Parkinson's** and exposure to pesticides in particular **Paraquat**^[9]
- 2 Also some evidence of association between **Parkinson's** and **DJ-1** gene^[10]
- 3 There also appears to be some evidence in fruit flies of **DJ-1 - Paraquat interaction**^[11]

GEI example, Parkinson's, exposure to pesticides and the DJ-1 gene

Based on work with P.Vineis and V.Gallo online at www1.imperial.ac.uk/medicine/about/divisions/ephpc/eph/eph_publications/techreports/

Story

- 1 Some evidence of association between **Parkinson's** and exposure to pesticides in particular **Paraquat**^[9]
- 2 Also some evidence of association between **Parkinson's** and **DJ-1** gene^[10]
- 3 There also appears to be some evidence in fruit flies of **DJ-1 - Paraquat interaction**^[11]

Question: Is there evidence of a similar interaction in humans?

Ottman/Khoury GEI models

- 1 Both Ottman^[12] and Khoury^[13] have put forward models for GEI.
- 2 Use models to try and understand type of interaction

Ottman/Khoury GEI models

- 1 Both Ottman^[12] and Khoury^[13] have put forward models for GEI.
- 2 Use models to try and understand type of interaction

Notation

- 1 gene: $DJ - 1 = 1$ mutation $DJ - 1 = 0$ normal
- 2 exposure to Paraquat: $P = 1$ exposed $P = 0$ unexposed
- 3 Parkinson's: $Y = 1$ presence $Y = 0$ absence
- 4 For now assume no confounders between P and Y
- 5 And independence of P and $DJ - 1$ i.e. $P \perp\!\!\!\perp DJ - 1$.

Interaction

- 1 To cater for idea of an interaction
- 2 introduce an **interaction variable** I

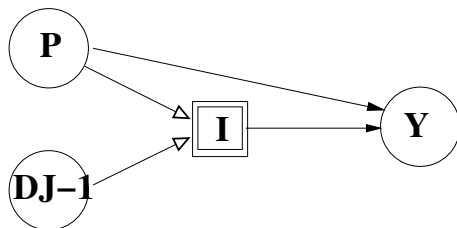
- 1 To cater for idea of an interaction
- 2 introduce an **interaction variable** I
- 3 I is such that
 - 1 I is binary - open/closed
 - 2 acts like a switch
 - 3 is different for each interaction

Interaction

- 1 To cater for idea of an interaction
- 2 introduce an **interaction variable** I
- 3 I is such that
 - 1 I is binary - open/closed
 - 2 acts like a switch
 - 3 is different for each interaction
- 4 Similar to concept of sufficient component causes^[14]
- 5 but simpler and “looser”
- 6 based on decision theoretic framework

Model 1:

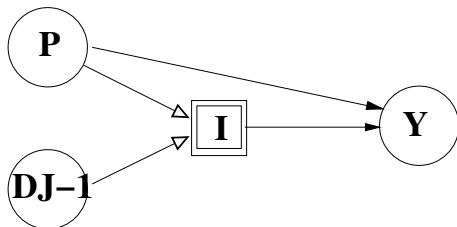
Genotype exacerbates effect of exposure, but there is no effect of genotype on unexposed.



- 1 I depends **only on exposure** $I \equiv P$
- 2 When $P = 1$ then I is turned “on” allowing interaction

Model 1:

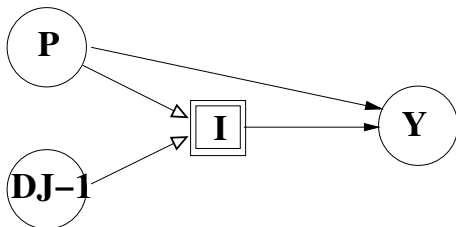
Genotype exacerbates effect of exposure, but there is no effect of genotype on unexposed.



- 1 **I** depends **only on exposure** $I \equiv P$
- 2 When $P = 1$ then **I** is turned “on” allowing interaction
- 3 Does this reflect evidence?

Model 1:

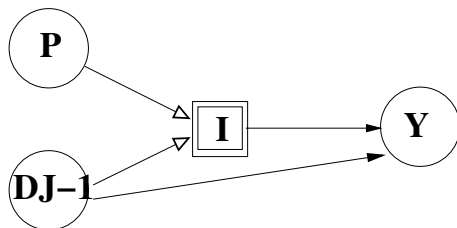
Genotype exacerbates effect of exposure, but there is no effect of genotype on unexposed.



- 1 I depends **only on exposure** $I \equiv P$
- 2 When $P = 1$ then I is turned “on” allowing interaction
- 3 Does this reflect evidence?
- 4 No - as families with $DJ - 1$ more likely to get Parkinson's irrespective of exposure to P

Model 2:

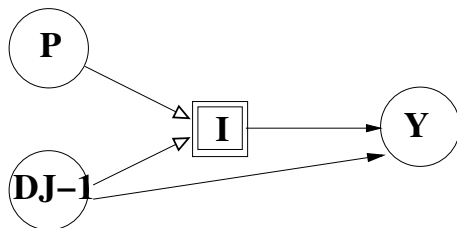
Exposure exacerbates the effect of genotype but has no effect on persons with low-risk genotype.



- 1 I depends **only on genotype** $I \equiv Dj - 1$
- 2 When $DJ - 1 = 1$ then I is turned “on” allowing interaction

Model 2:

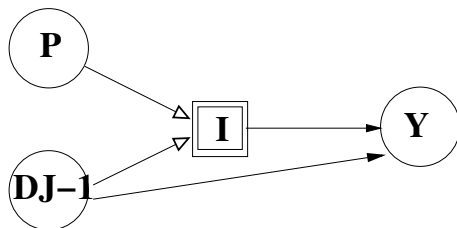
Exposure exacerbates the effect of genotype but has no effect on persons with low-risk genotype.



- 1 I depends **only on genotype** $I \equiv Dj - 1$
- 2 When $DJ - 1 = 1$ then I is turned “on” allowing interaction
- 3 Does this reflect evidence?

Model 2:

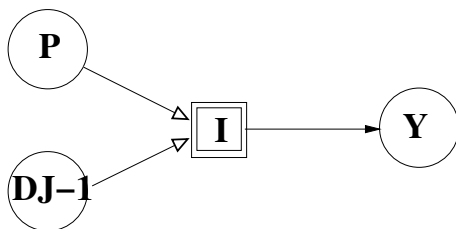
Exposure exacerbates the effect of genotype but has no effect on persons with low-risk genotype.



- 1 I depends **only on genotype** $I \equiv Dj - 1$
- 2 When $DJ - 1 = 1$ then I is turned “on” allowing interaction
- 3 Does this reflect evidence?
- 4 No - as people exposed to P more likely to get Parkinson's even when they do not have $DJ - 1$ mutation

Model 3:

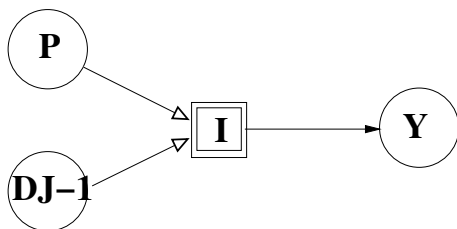
Both exposure and genotype are required to increase risk



- 1 I depends on **both** genotype and exposure
- 2 So if $DJ - 1 = 1$ **and** $P = 1$ then I is turned “on” otherwise “off”

Model 3:

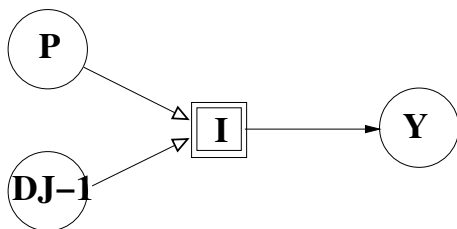
Both exposure and genotype are required to increase risk



- 1 I depends on **both** genotype and exposure
- 2 So if $DJ - 1 = 1$ **and** $P = 1$ then I is turned “on” otherwise “off”
- 3 Does this reflect evidence?

Model 3:

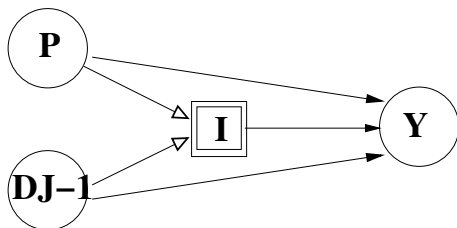
Both exposure and genotype are required to increase risk



- 1 I depends on **both** genotype and exposure
- 2 So if $DJ - 1 = 1$ **and** $P = 1$ then I is turned “on” otherwise “off”
- 3 Does this reflect evidence?
- 4 No - follows from comments about previous models
- 5 Very rarely the case e.g. **favism**

Model 4:

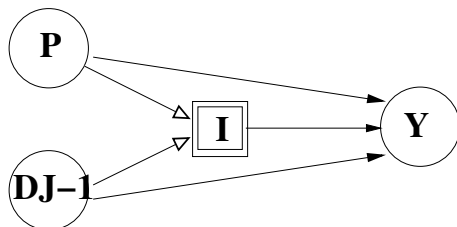
Exposure and genotype both have some effect but together they have a worse (better) effect



- 1 I depends depends on **both** genotype and exposure
- 2 So I is only “off” if $DJ - 1 = 0$ **or** $P = 0$
- 3 Does this reflect evidence?

Model 4:

Exposure and genotype both have some effect but together they have a worse (better) effect

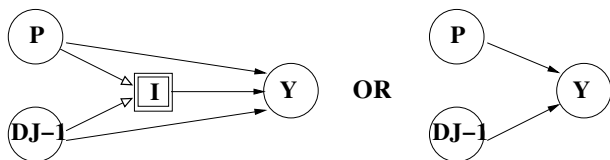


- 1 I depends depends on **both** genotype and exposure
- 2 So I is only “off” if $DJ - 1 = 0$ **or** $P = 0$
- 3 Does this reflect evidence?
- 4 Yes it does

Problem with Interaction models

E.g. Model 4

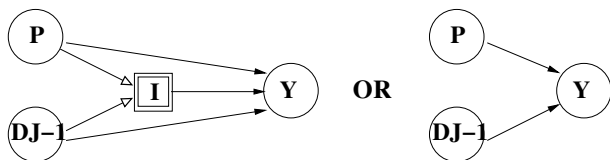
- 1 How to distinguish this model from just P and $DJ-1$ each have effects?



Problem with Interaction models

E.g. Model 4

- 1 How to distinguish this model from just P and $DJ-1$ each have effects?



- 2 Could use a large study
- 3 Could also be resolved using experiment
- 4 Not ethical/possible for humans but it has been done for fruit flies

GEI - fruit flies

Exposure and genotype both have some effect but together they have a worse (better) effect

- Experiments where flies have been exposed to paraquat
 $F_P = 1$

GEI - fruit flies

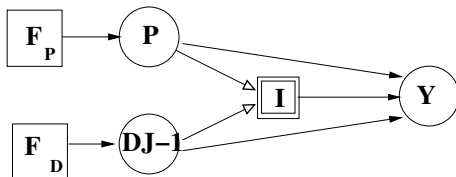
Exposure and genotype both have some effect but together they have a worse (better) effect

- Experiments where flies have been exposed to paraquat
 $F_P = 1$
- and had there $DJ - 1$ gene knocked out and mutated
 $F_D = 1, 0$
- And it would appear there is an interaction

GEI - fruit flies

Exposure and genotype both have some effect but together they have a worse (better) effect

- Experiments where flies have been exposed to paraquat $F_P = 1$
- and had there $DJ - 1$ gene knocked out and mutated $F_D = 1, 0$
- And it would appear there is an interaction
- In this case **can distinguish between interaction/no interaction model**



Conclusion

- 1 DAGs are useful to visualise complex problems

Conclusion

- 1 DAGs are useful to visualise complex problems
- 2 But can also be looked at as formal tools as they correspond to CI's
- 3 When data are observational DAGs can only express association,

Conclusion

- 1 DAGs are useful to visualise complex problems
- 2 But can also be looked at as formal tools as they correspond to CI's
- 3 When data are observational DAGs can only express association, causation must be introduced based on other knowledge

Conclusion

- 1 DAGs are useful to visualise complex problems
- 2 But can also be looked at as formal tools as they correspond to CI's
- 3 When data are observational DAGs can only express association, causation must be introduced based on other knowledge
- 4 DT is one way of doing introducing causality via intervention

Conclusion

- 1 DAGs are useful to visualise complex problems
- 2 But can also be looked at as formal tools as they correspond to CI's
- 3 When data are observational DAGs can only express association, causation must be introduced based on other knowledge
- 4 DT is one way of doing introducing causality via intervention

Further work

- 1 Generally use more DAGs in Epidemiology
- 2 Interaction models too simple
- 3 Look at more complex interaction models where G and E are associated
- 4 where interaction is not deterministic but stochastic and depends on other unobserved factors not in the DAG

Bibliography

1. Dawid AP 1979 J Roy Statist Soc B. 41:1-31.
2. Nissen S. et al, 2005, NEJM, 352(1)
3. Ridker P. et al, 2002, NEJM, 347(20)
4. Ridker P. et al, 2005, NEJM, 352(1)
5. Dawid AP, 2002, Intern. Statist. Rev. 70
6. Cartwright N. 1994 Oxford Monographs
7. Davey SG, Ebrahim S. 2003 Int J Epi.,32(1)
8. Edwin M. et al 2006, Epidemiology, 17(3)
9. Brown TP,et al. 2006 Environ Health Perspect.,114(2)
- 10.Clements CM et al. 2006 Proc Natl Acad Sci U S A.103(41)
- 11.Meulener M et al. Current Bio,15
- 12.Ottman R. Genet Epidemiol. 1990,7(3)
- 13.Khoury M. et al. OUP. 1993
- 14.VanderWeele T., Robins J. 2007,Epidemiology,18(3)