Directed Acyclic Graphs (DAGs)
And Causal Structures of Systematic Biases

UCSF
University of California
San Francisco

Cindy Leung, ScD, MPH
Center for Health and Community

1/12/2017

Highlights from Week #1

- **Theories of causal inference**
  - Inductive
  - Deductive
  - Other (refutationist, Bayesian)
- **Causal models (heuristics)**
  - A. Bradford Hill's causal guidelines
  - Sufficient-component cause model ('causal pies')
  - Counterfactual model ('potential outcomes')
  - Graphic models (DAGs)
  - Agent-based modeling (multi-dimensional causal matrix)
- **Practice of causal inference**

Highlights from Week #1

- **Sufficient-component cause model**
  - Advantages:
    - Useful for examining multiple risk factors that act in concert
    - Can show interactions among causes
  - Disadvantages:
    - Omits discussion of origins of causes
    - Focuses on proximal causes
    - Ignores induction period
    - Does not address indirect effects
    - Focuses on causes of diseases in individuals or groups, but not in populations
    - Does not consider factors that control distributions of risk factors
    - Ignores dynamic non-linear relations

Highlights from Week #1

- **Counterfactual model**
  - Can be applied at the individual or population level
  - Specifies what would happen to individuals or populations under alternative patterns of exposure
  - Forces researchers to think about operational definitions of cases and controls, sampling schemes, and other important design questions
  - One of the two conditions in the definitions of the effect measures must be contrary to the fact – exposures or treatment vs. a reference condition
  - More on this in this lecture!

Highlights from Week #1

- **Theories of causal inference**
  - Inductive
  - Deductive
  - Other (refutationist, Bayesian)
- **Causal models (heuristics)**
  - A. Bradford Hill's causal guidelines
  - Sufficient-component cause model ('causal pies')
  - Counterfactual model ('potential outcomes')
  - Graphic models (DAGs)
  - Agent-based modeling (multi-dimensional causal matrix)
- **Practice of causal inference**

Causal Diagrams for Epidemiologic Research

Sander Greenland,[1] Judea Pearl,[2] and James M. Robins[3]

Causal diagrams have a long history of informal use and, more recently, have undergone formal development for applications in expert systems and robotics. We provide an introduction to these developments and their use in epidemiologic research. Causal diagrams can provide a starting point for identifying variables that must be measured and controlled to obtain unconfounded effect estimates. They also provide a method for critical evaluation of traditional epidemiologic criteria for confounding. In particular, they reveal certain heretofore unnoticed shortcomings of those criteria when used in considering multiple potential confounders. We show how to modify the traditional criteria to correct those shortcomings. (Epidemiology 1999;10:37–48)

Keywords: bias, causation, confounding, epidemiologic methods, graphical methods, observational studies.

Summarization of causal links via graphs or diagrams has long been used as an informal aid to causal analysis. Causal graphs in the form of path diagrams are an integral component of path analysis[1] and structural equations modeling.[2] In more recent times, the theory of directed acyclic graphs (DAGs) has been extended to application in expert-systems research.[3,4] In these applications, there is a pressing need for valid formal rules that allow an automated system or robot to deduce correctly the presence or absence of causal links given correct background information and new data. The outgrowth of this research has been the development of a formal theory for evaluating causal effects using the language of causal diagrams.[5,6] Unlike path analysis and structural-equations modeling, this theory does not require parametric assumptions such as linearity.

The theory of causal graphs is equivalent to the G-computation theory of Robins.[7–9] It has a benefit, however, of providing a compact graphical as well as algebraic formulation of assumptions and results, which may be easier for the general reader to comprehend. In addition, it provides a novel perspective on traditional epidemiologic criteria for confounder identification.

We here provide a brief introduction to the theory of causal diagrams based on DAGs.[5,6] We pay special attention to its relation to nongraphical epidemiologic treatments of confounding.[10–13] We show how diagrams can serve as a visual yet logically rigorous aid for summarizing assumptions about a problem and for identifying variables that must be measured and controlled to obtain unconfounded effect estimates given those assumptions. Thus, use of such graphs can aid in planning of data collection and analysis, in communication of results, and in avoiding subtle pitfalls of confounder selection.

Except where noted otherwise, the present paper will deal only with relations among variables in a given source population; that is, we will deal only with structural (systematic) relations among the underlying variables of interest, so that issues of measurement error and random variation will not arise. We will also not present proofs of results, but we will give references in which proofs can be found.

A Rationale for Graphs

Any deduction about a causal relation must start from

"Causal diagrams can provide a starting point for identifying variables that must be measured and controlled to obtain unconfounded effect estimates. They also provide a method for critical evaluation of traditional epidemiologic criteria for confounding."

Learning Objectives

- Understand rules for causal DAGs
- Examine DAGs depicting confounding bias
- Examine DAGs depicting selection bias
- Apply DAGs to real-life studies
- Discuss current controversies around DAGs

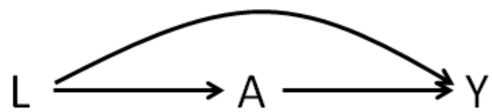
Learning Objectives

- **Understand rules for causal DAGs**
- Examine DAGs depicting confounding bias
- Examine DAGs depicting selection bias
- Apply DAGs to real-life studies
- Discuss current controversies around DAGs

Causal DAGs

- DAGs are used to:
  - Think conceptually about causal inference
  - Identify sources of bias in a study
  - Identify approaches to study design and analysis to eliminate systematic bias and to help identify average causal effects

- **D: Directed**
- A: Acyclic
- G: Graphs

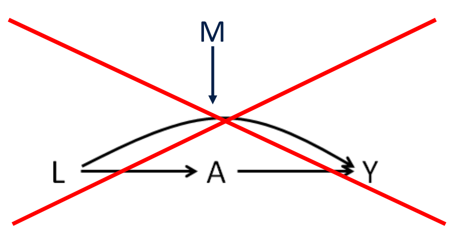


1/12/2017

Causal DAGs

- DAGs are used to:
  - Think conceptually about causal inference
  - Identify sources of bias in a study
  - Identify approaches to study design and analysis to eliminate systematic bias and to help identify average causal effects

- **D: Directed**
- A: Acyclic
- G: Graphs

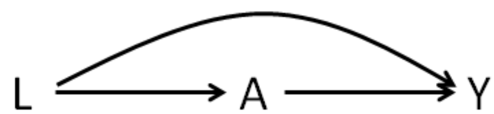


1/12/2017

Causal DAGs

- DAGs are used to:
  - Think conceptually about causal inference
  - Identify sources of bias in a study
  - Identify approaches to study design and analysis to eliminate systematic bias and to help identify average causal effects

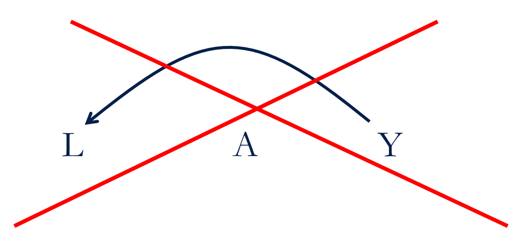
- D: Directed
- **A: Acyclic**
- G: Graphs



Causal DAGs

- DAGs are used to:
  - Think conceptually about causal inference
  - Identify sources of bias in a study
  - Identify approaches to study design and analysis to eliminate systematic bias and to help identify average causal effects

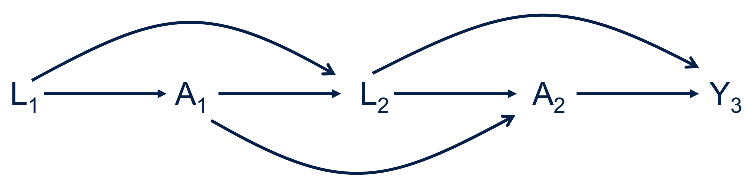
- D: Directed
- **A: Acyclic**
- G: Graphs



Causal DAGs

- DAGs are used to:
  - Think conceptually about causal inference
  - Identify sources of bias in a study
  - Identify approaches to study design and analysis to eliminate systematic bias and to help identify average causal effects

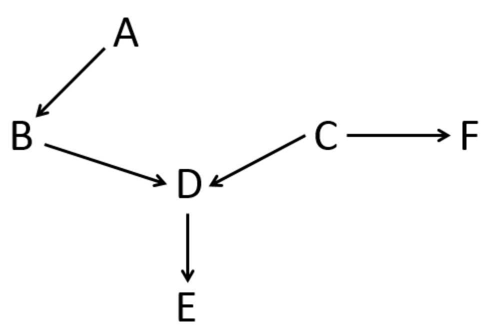
- D: Directed
- **A: Acyclic**
- G: Graphs



UCSF

Vocabulary for Causal DAGs

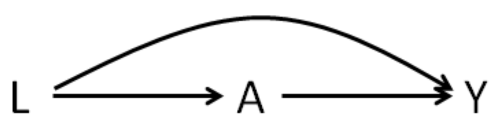
- **Nodes**: variables
- **Direct effect**: A directly affects Y if there is one arrow from A to Y
- **Indirect effect**: A indirectly affects Y if there is an intermediate between A to Y

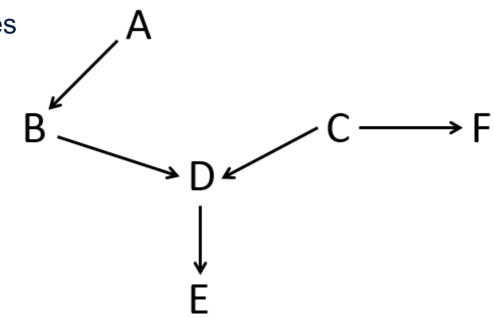


UCSF

Vocabulary for Causal DAGs

- **Descendants**: variables directly affected by your variable of interest
- **Ancestors**: variables that directly affect your variable of interest
- **Path**: Any sequence of arrows that connect two variables
- **Collider**: A variable directly affected by two variables

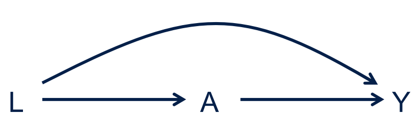




Examples of causal DAGs

- Marginally randomized experiment

- Conditionally randomized experiment



Drawing causal DAGs

- Identify all random variables of interest/ common causes of exposure and outcome
  - A DAG is causal only if all common causes are shown
  - Variables not common causes do not need to be shown
- Variables drawn in temporal order
- Draw arrows in direction of causation
- Complete DAGs do not make any assumptions
- Missing arrows convey expert knowledge about the association of interest

$$\begin{tikzcd} L \arrow[r] \arrow[rr, bend left] & A \arrow[r] & Y \end{tikzcd}$$

$$\begin{tikzcd} L \arrow[r] & A \arrow[r] & Y \end{tikzcd}$$

Two variables are associated if…

1) **Direct causal effect**

$$A \longrightarrow Y$$

- A: high red meat intake Y: colorectal cancer
- B: heme iron

Two variables are associated if…

2) Shared common causes

$$\begin{tikzcd} L \arrow[r] \arrow[rr, bend left] & A & Y \end{tikzcd}$$

- A: high white potatoes intake
- L: high red meat intake

Y: colorectal cancer

Two variables are associated if…

3) Conditioning on a common effect

- Sometimes called "collider bias"

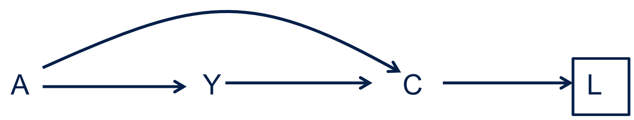
$$\begin{tikzcd} A \arrow[rr, bend left] & Y \arrow[r] & \boxed{C} \end{tikzcd}$$

- A: raining
- C: wet grass

Y: sprinkler turned on

Two variables are associated if…

3.1) Conditioning on the descendant of a common effect



- A: raining
- C: wet grass

Y: sprinkler turned on

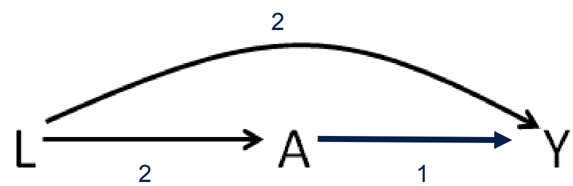
L: slipped on the grass

D-separation terminology

- Set of graphical rules to decide whether two variables are:
  - Independent (d-separated)
  - Associated (d-connected)
- Path is any arrow-based route between two variables in the graph
  - Not dependent on arrow directions
  - Paths can be blocked or open according to a specific set of rules
- Conditioning
  - Comparing the levels of one variable with another
  - Study design phase: restriction, matching
  - Study analysis phase: stratification, adjustment via regression

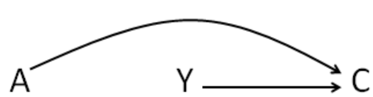
D-separation rules

- Each path goes between the exposure and the outcome
  - Disregard the arrow directions
  - Do not pass through a node more than once

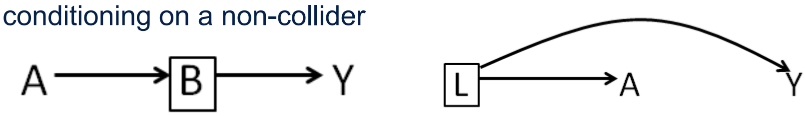


D-separation rules

- A path is blocked:
  - If there is a collider

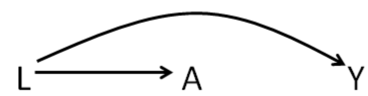


  - When conditioning on a non-collider

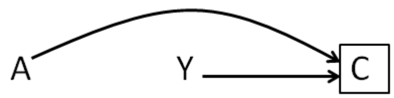


D-separation rules

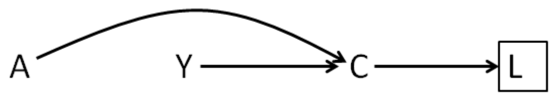
- A path is open:
  - When no non-colliders have been conditioned



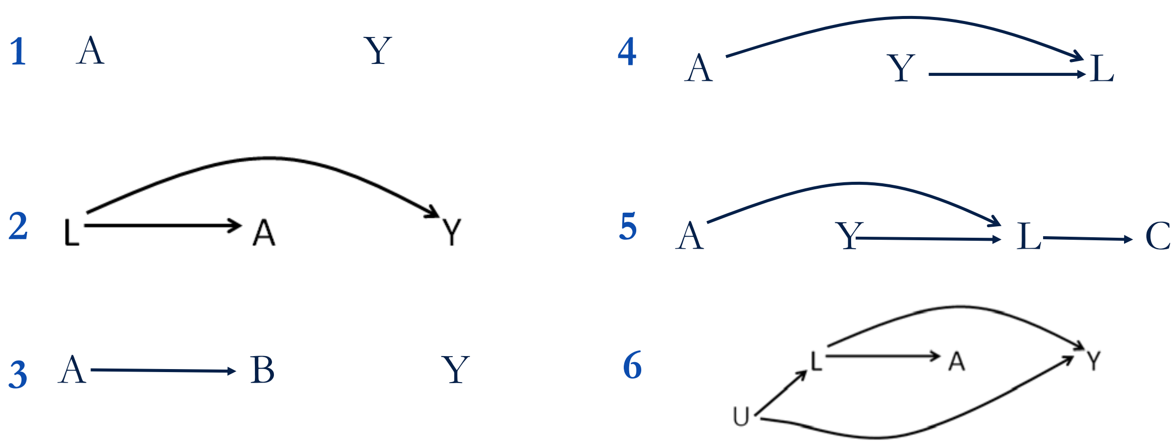
  - When conditioning on a collider



  - When conditioning on a descendent of a collider



Sample DAGs – test your knowledge



Learning Objectives

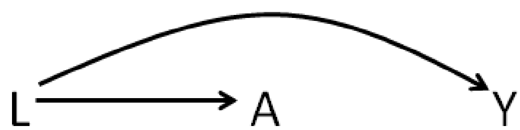
- Understand rules for causal DAGs
- **Examine DAGs depicting confounding bias**
- Examine DAGs depicting selection bias
- Apply DAGs to real-life studies
- Discuss current controversies around DAGs

Confounders

- Standard definition
  - L is associated with A
  - L is associated with Y independent of A
  - L is not on the causal pathway between A and Y
- Collapsibility definition
  - L is a confounder if conditional on L, the association between A and Y is different from the unconditional association
- Causal definition
  - A variable that can be used to block a backdoor path

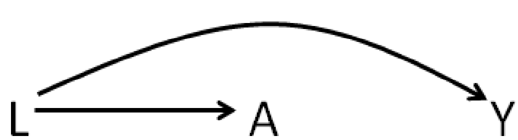
Confounding

- Bias due to common causes
  - The exposed and unexposed are not exchangeable

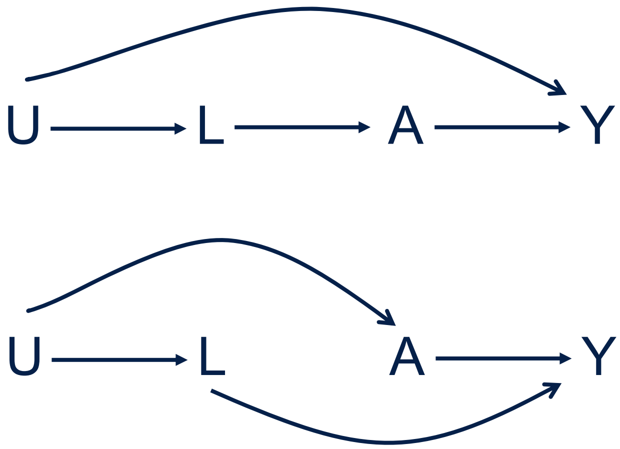


Confounding : an example

- L: Education
- A: Aspirin use
- Y: Heart disease



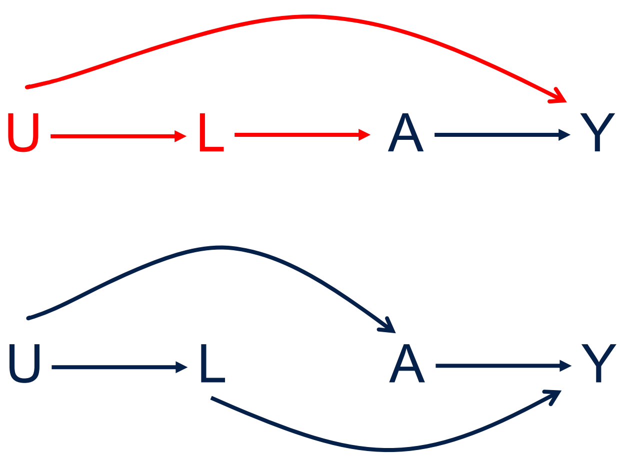
Confounder?



- Is confounding present?
- Which variable(s) encompass the standard definition of a confounder?
- Which variable(s) can be used to block the backdoor path between A and Y?

1/12/2017

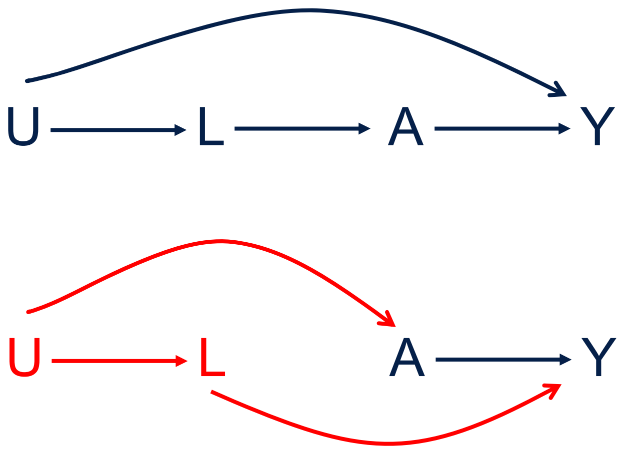
Confounder?



- Is confounding present?
- Which variable(s) encompass the standard definition of a confounder?
- Which variable(s) can be used to block the backdoor path between A and Y?

1/12/2017

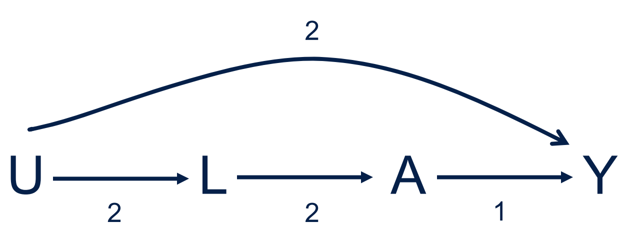
Confounder?



- Is confounding present?
- Which variable(s) encompass the standard definition of a confounder?
- Which variable(s) can be used to block the backdoor path between A and Y?

Another example

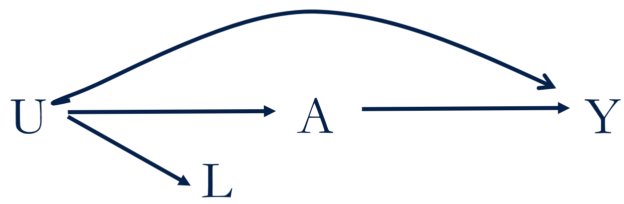
- U: Genetic factor
- L: Family history of malformation
- A: Multivitamin use during pregnancy
- Y: Congenital malformation



Hernan et al., Causal knowledge as a Prerequisite for Confounding Epidemiology: An Application to Birth Defects Epidemiology. AJE 2002.

Proxies for confounders

- U: Socioeconomic status
- L: Education, income
- A: Poor diet
- Y: Obesity



Unmeasured confounding exists even after conditioning on L.

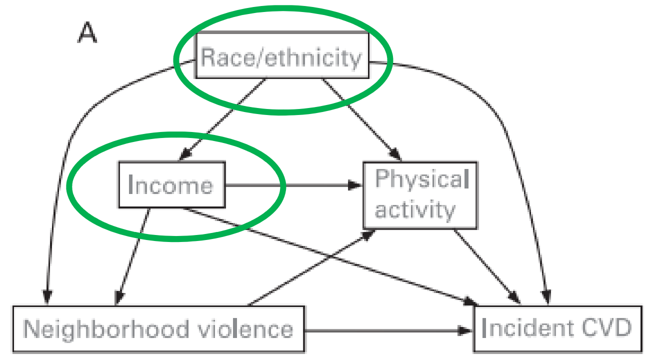
The strength of the association between U and L can indicate degree of unmeasured confounding.

So what?

- We need expert knowledge to determine if we should adjust for a variable
  - Statistical criteria are insufficient

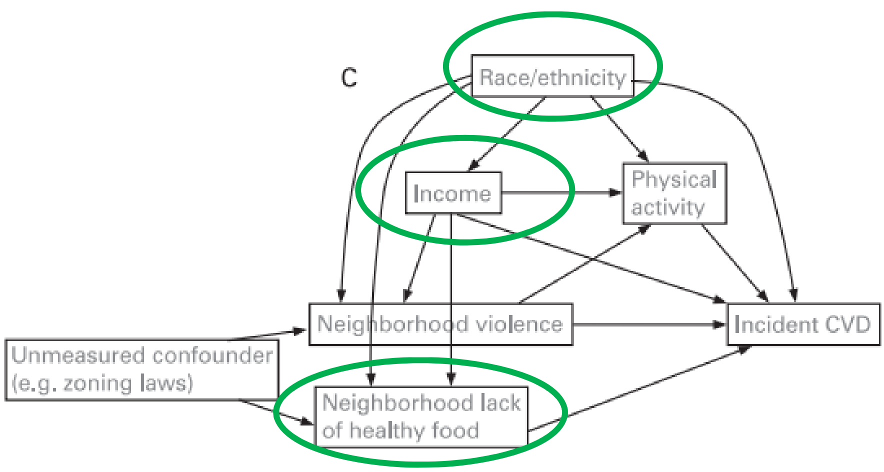
- Causal DAGs allow us to integrate expert knowledge into an analysis to identify true confounders

Neighborhood violence and CVD



Fleischer NL, Diez Roux AV. Using directed acyclic graphs to guide analyses of neighbourhood effects: an introduction. JECH 2009.

Neighborhood violence and CVD

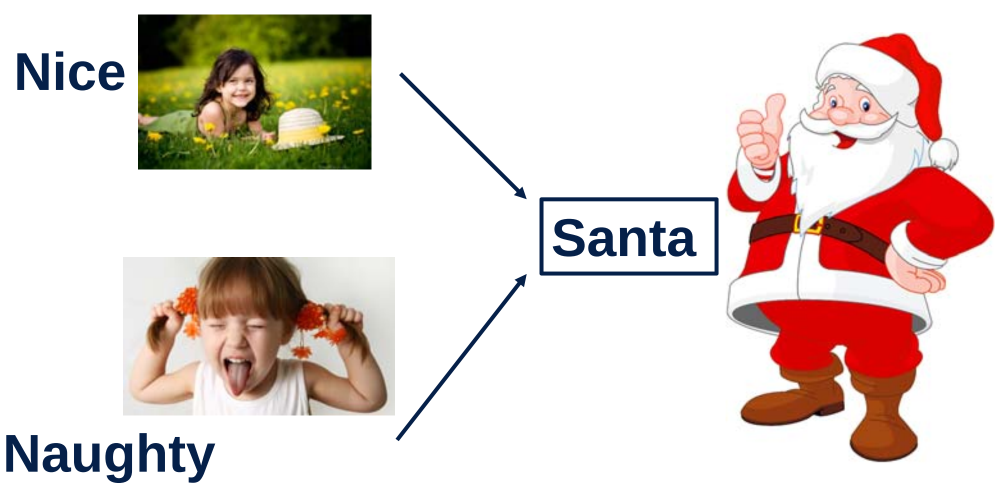


Fleischer NL, Diez Roux AV. Using directed acylic graphs to guide analyses of neighbourhood effects: an introduction. JECH 2009.

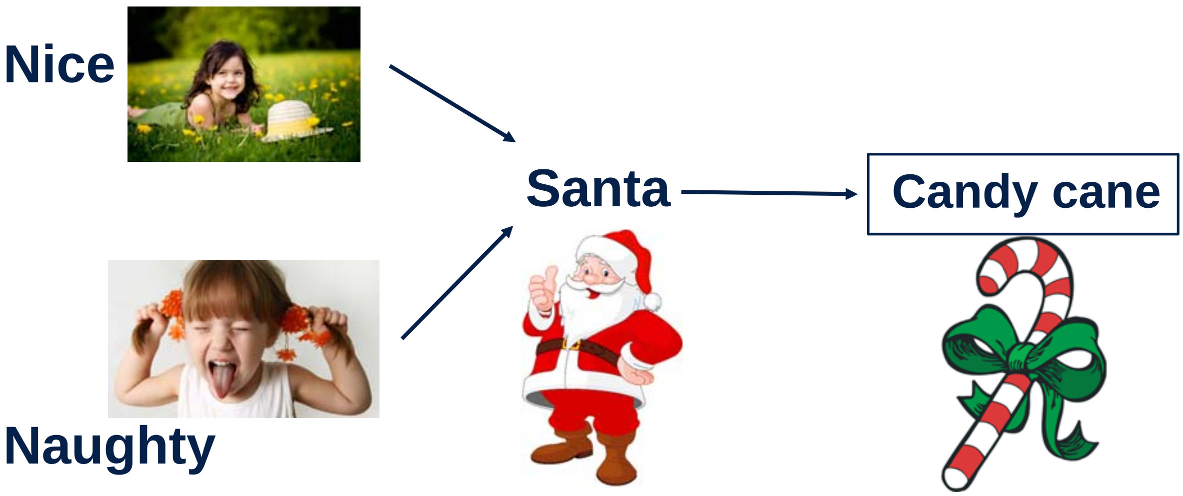
Learning Objectives

- Understand rules for causal DAGs
- Examine DAGs depicting confounding bias
- **Examine DAGs depicting selection bias**
- Apply DAGs to real-life studies
- Discuss current controversies around DAGs

Selection/ Collider bias



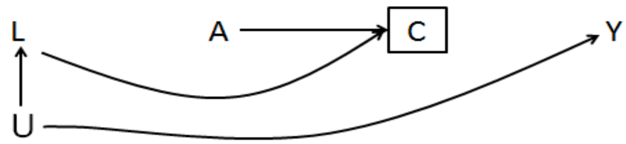
Selection/ Collider bias



Selection bias

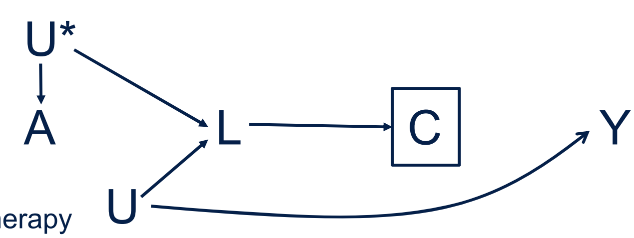
- Systematic bias due to the selection of subjects into the study or the analysis
  - Conditioning on a common effect
    - Survival
    - Diagnosis with disease
    - Selection into a study
    - Complete data

Differential loss to Follow-Up



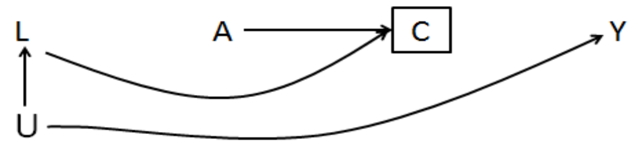
- A: Antiretroviral therapy
- Y: AIDS
- C: Lost during follow-up
- L: CD4 count
- U: Immunologic status

Differential loss to Follow-Up



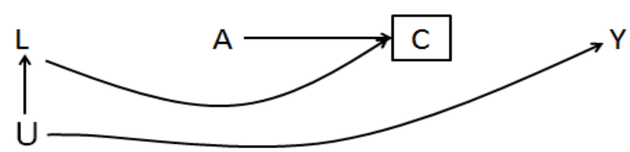
- A: Antiretroviral therapy
- Y: AIDS
- C: Lost during follow-up
- L: CD4 count
- U: Immunologic status
- U*: Health behaviors

Volunteer/ self-selection



- A: Smoking
- Y: Heart disease
- C: Volunteer to participate
- L: Awareness of heart disease
- U: Family history of heart disease

Healthy Worker Bias



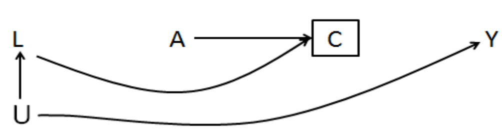
- A: Occupational exposure
- Y: Death
- C: Being at work
- L: Self-reported health status
- U: Underlying health status

What do these DAGs have in common?

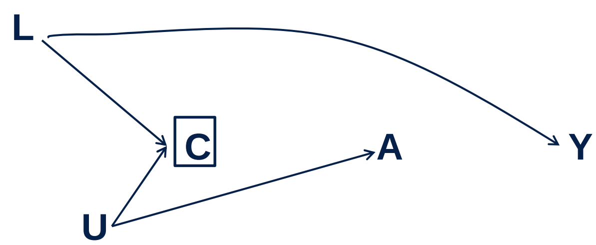
- All DAGs represent conditioning on a common effect of:
  - The exposure and the outcome, or
  - A cause of the exposure and a cause of the outcome
  - Not controlled by the investigator (usually)
- The conditioning of the common effect opens an alternate path from A to Y that is not the direct effect of interest
  - Unlike confounding, not a backdoor path
  - Perhaps a “frontdoor path”?

Control of selection bias

- In the design phase:
  - Sample controls in a manner that ensure that they represent the exposure distribution in the population
- In the analytic phase:
  - Stratification
  - IPW

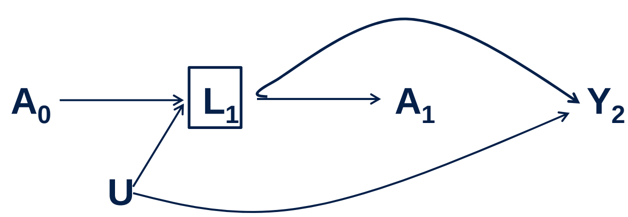


Confounding or Selection Bias?



- A: Physical Activity
- Y: Heart disease
- C: Being a firefighter
- L: Parental SES
- U: Attraction to activities that involve physical activity

Confounding or Selection Bias?



- A: Aspirin use
- Y: Stroke
- L: Gastrointestinal bleeding
- U: Underlying health status

Learning Objectives

- Understand rules for causal DAGs
- Examine DAGs depicting confounding bias
- Examine DAGs depicting selection bias
- **Apply DAGs to real-life studies**
- Discuss current controversies around DAGs

Drawing a DAG

- Identify all the variables of interest
  - Exposure, outcome, common causes, modifying variables, common effects, selection factors, etc.
  - Draw on the expert knowledge of your field and research team
  - Draw your DAG before running any analyses (not data driven) and preferably BEFORE you collect any data
- List your variables in temporal order (so arrows point to the right)
- Find all paths between the exposure and the outcome
  - Distinguish between causal and non-causal paths
  - Distinguish between open and closed paths
- Determine the minimally sufficient set of variables to condition on so that:
  - All causal paths will be open
  - All non-causal paths will be blocked

Try it!

You are part of a team evaluating the effects of a **mindful eating intervention** on **excessive gestational weight gain**. Women were not randomized; however, the intervention and control group differed only on parity.

Your PI suggests excluding women who developed gestational diabetes (measured at 24 weeks gestation) from the main analysis, even though they had complete follow-up through delivery.

1. Use a DAG to guide your decisions on which variables should be included in the main analysis of intervention effect on gestational weight gain.
2. Should cases of gestational diabetes be excluded from the analysis? Why/why not?

MAMAS intervention DAG

Mindfulness intervention

Gestational weight gain

MAMAS intervention DAG

Prepregnancy BMI

Maternal age

Household income

Mindfulness intervention

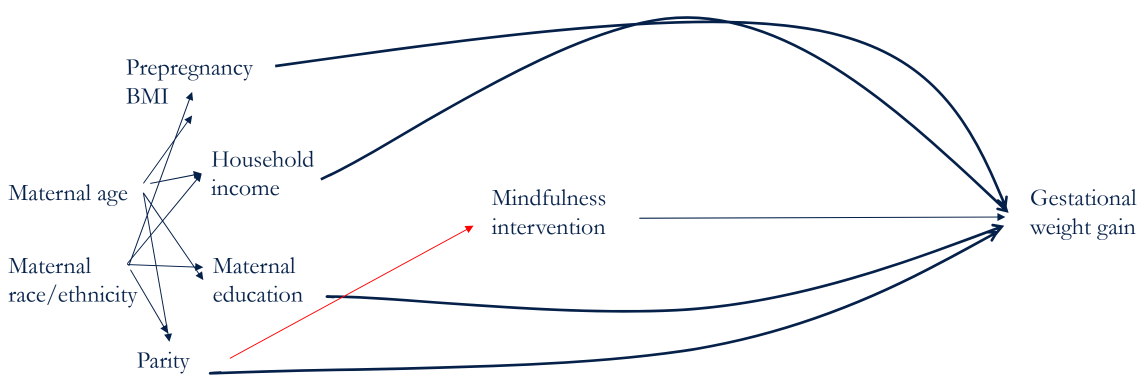
Gestational weight gain

Maternal race/ethnicity

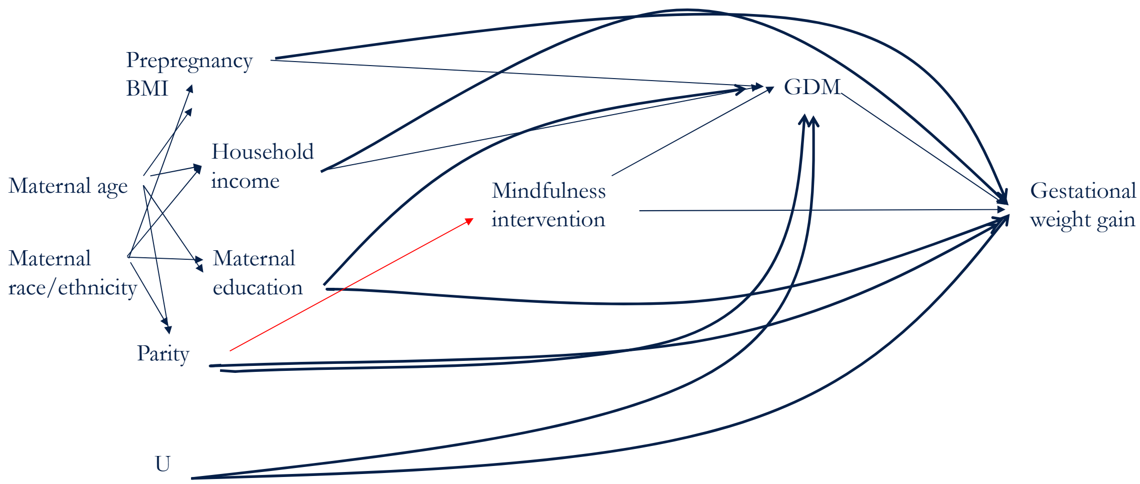
Maternal education

Parity

MAMAS intervention DAG

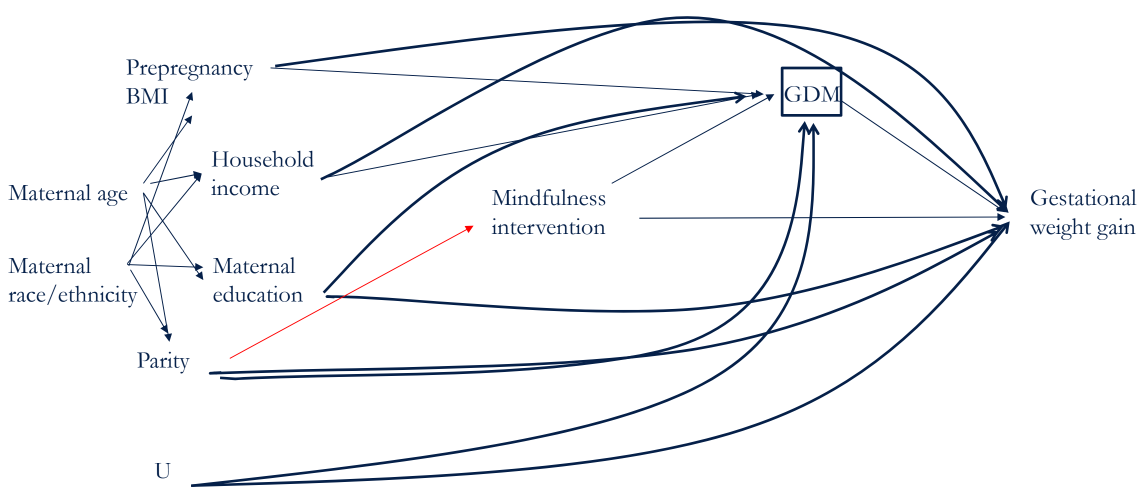


MAMAS intervention DAG



1/12/2017

MAMAS intervention DAG



1/12/2017

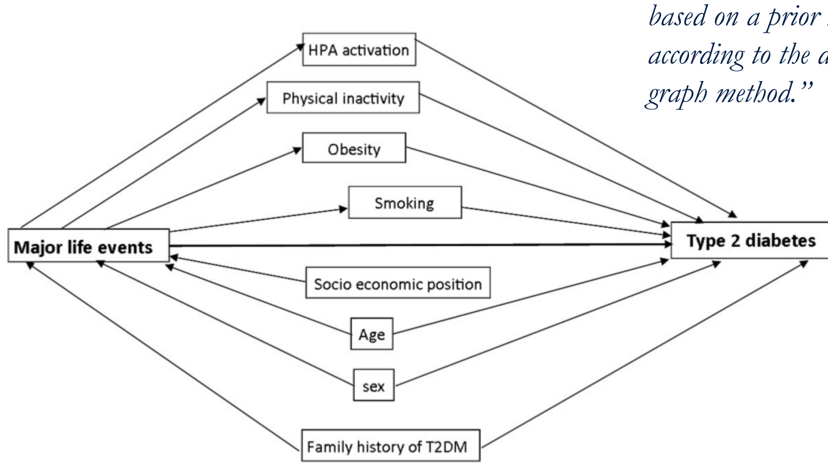
MAMAS intervention DAG

Should cases of gestational diabetes be excluded from the analysis? Why/why not?

- No, excluding GDM (conditioning on GDM, a collider) will open a non-causal path between A and Y through GDM and unmeasured confounders of GDM and GWG that will induce a bias

Other examples

"Confounders were identified based on a prior knowledge and according to the directed acyclic graph method."



Pedersen JM et al. Accumulation of major life events in childhood and adult life and risk of type 2 diabetes mellitus. PLoS One. 2015.

Learning Objectives

- Understand rules for causal DAGs
- Examine DAGs depicting confounding bias
- Examine DAGs depicting selection bias
- Apply DAGs to real-life studies
- **Discuss current controversies around DAGs**

DAGs vs. Change-in Estimate Method

1. **DAG full model:** only covariates identified as confounders in the DAG were included
2. **DAG gold-standard change-in-estimate procedure**: start with initial DAG full model and remove 1 covariate at a time with the smallest change in OR compared to OR in the full model
3. **DAG gold-standard change-in-estimate procedure with consideration of precision:** compare precision of model estimates of #1 to #2, and select #2 only if it has smaller standard error than #1
4. **DAG-stepwise change-in-estimate procedure**: start with initial DAG full model and remove 1 covariate at a time with the smallest change in OR compared to the previous step

Weng et al. Methods of Covariate Selection: Directed Acylic Graphs and the Change-in-Estimate Procedure. AJE 2009.

DAGs vs. Change-in Estimate Method

Table 3. Scenarios Based on Correctly Specified Directed Acyclic Graphs Representing Confounders of Various Strengths[a]

Measures by Scenario[b]	DAG	DAG-GS CE	DAG-GS P	DAG-S CE
Standard error				
1	0.173	0.177	0.177	0.188
2	0.164	0.167	0.167	0.179
3	0.176	0.177	0.177	0.184
Bias				
1	0.007	0.057	0.057	0.077
2	−0.006	0.058	0.058	0.119
3	0.003	0.072	0.072	0.159

Scenarios 1-3 contain confounders of decreasing strengths

Weng et al. Methods of Covariate Selection: Directed Acylic Graphs and the Change-in-Estimate Procedure. AJE 2009.

UCSF

DAGs vs. Change-in Estimate Method

"For modeling binary outcomes in a case-control study, the authors recommend construction of a "conservative" DAG, determination of all potential confounders, and then change-in-estimate procedures to simplify this full model.... The selection of final model should be based on changes in precision.

Adopt the reduced model if its standard error is substantially smaller; otherwise, the full DAG-based model is appropriate."

Weng et al. Methods of Covariate Selection: Directed Acylic Graphs and the Change-in-Estimate Procedure. AJE 2009.

UCSF

Some Limitations of DAGs

- DAGs don't provide info on the strength or direction of the confounding or selection bias
  - Can use + and – on arrows with hypothesized directions
  - Only works for binary variables and for direction of confounding
- DAGs don't provide info on how bias will affect precision of estimates
- DAGs do not easily show effect modification
- DAGs are not well-suited for multiple exposures, outcomes or multi-level exposures.

UCSF

"In our view, no one causal approach should drive the questions asked or delimit what counts as useful evidence. Robust causal inference instead comprises a complex narrative, created by scientists appraising, from diverse perspectives, different strands of evidence produced by myriad methods. DAGs can of course be useful, but should not alone wag the causal tale."

IJEA

International Journal of Epidemiology, 2016, 1–22
doi: 10.1093/ije/dyw114
Original article

Original article

The tale wagged by the DAG: broadening the scope of causal inference and explanation for epidemiology

Nancy Krieger[1,*] and George Davey Smith[2]

[1]Department of Social and Behavioral Sciences, Harvard T.H. Chan School of Public Health, Boston, MA, USA and, [2]MRC Integrative Epidemiology Unit at the University of Bristol, Bristol, UK

*Corresponding author. Department of Social and Behavioral Sciences, Kresge 717, Harvard T.H. Chan School of Public Health, 677 Huntington Avenue, Boston, MA 02115, USA. E-mail: nkrieger@hsph.harvard.edu

Accepted 13 April 2016

Abstract

'Causal inference', in 21st century epidemiology, has notably come to stand for a specific approach, one focused primarily on counterfactual and potential outcome reasoning and using particular representations, such as directed acyclic graphs (DAGs) and Bayesian causal nets. In this essay, we suggest that in epidemiology no one causal approach should drive the questions asked or delimit what counts as useful evidence. Robust causal inference instead comprises a complex narrative, created by scientists appraising, from diverse perspectives, different strands of evidence produced by myriad methods. DAGs can of course be useful, but should not alone wag the causal tale. To make our case, we first address key conceptual issues, after which we offer several concrete examples illustrating how the newly favoured methods, despite their strengths, can also: (i) limit who and what may be deemed a 'cause', thereby narrowing the scope of the field; and (ii) lead to erroneous causal inference, especially if key biological and social assump-

Method	Strengths in comparison with conventional observational analysis	Key assumptions
Observational studies	Not applicable	No unmeasured confounders; no measurement error in assessed confounders; correctly specified model (this assumption relates to all methods to a greater or lesser degree)
Cross-contextual comparisons	Will reveal context-specific confounding	No unmeasured confounding which (unlike the assessed confounders) is similar in magnitude between contexts and contributes substantially to the observed associations
Negative control studies	Reveals existence of potential unmeasured confounding.	Negative control is associated with confounders to the same extent as the exposure (or outcome) of interest is associated with these confounders
Within-sibship studies	Robust to fixed maternal effects that could confound the association	The important confounders do not change between pregnancies in a manner that is associated with change in maternal smoking behaviour
Children of twins	Between-MZ maternal twin pair analysis not subject to genetic confounding, or confounding by other factors that are shared between monozygotic twins. Comparison of between-MZ with between-DZ twin analyses allows estimation of extent of genetic confounding	No unmeasured confounding by factors that differ between twins
Mendelian randomization (MR)	no reverse causation	No pleiotropic effect of the genetic variants that influence the outcome independent of the exposure of interest
Non-genetic IVs	No systematic confounding	The instrumental variable does not relate to confounding factors and does not impact on the outcome except through the exposure of interest
Randomized controlled trials (RCTs)	Randomization leads to no systematic confounding	The intervention does not have effects except through changes in the exposure of interest

UCSF

"In our view, no one causal approach should drive the questions asked or delimit what counts as useful evidence. Robust causal inference instead comprises a complex narrative, created by scientists appraising, from diverse perspectives, different strands of evidence produced by myriad methods. DAGs can of course be useful, but should not alone wag the causal tale."

IEA

International Journal of Epidemiology, 2016, 1–22
doi: 10.1093/ije/dyw114
Original article

Original article

The tale wagged by the DAG: broadening the scope of causal inference and explanation for epidemiology

Nancy Krieger[1,*] and George Davey Smith[2]

[1]Department of Social and Behavioral Sciences, Harvard T.H. Chan School of Public Health, Boston, MA, USA and, [2]MRC Integrative Epidemiology Unit at the University of Bristol, Bristol, UK

*Corresponding author. Department of Social and Behavioral Sciences, Kresge 717, Harvard T.H. Chan School of Public Health, 677 Huntington Avenue, Boston, MA 02115, USA. E-mail: nkrieger@hsph.harvard.edu

Accepted 13 April 2016

Abstract

'Causal inference', in 21st century epidemiology, has notably come to stand for a specific approach, one focused primarily on counterfactual and potential outcome reasoning and using particular representations, such as directed acyclic graphs (DAGs) and Bayesian causal nets. In this essay, we suggest that in epidemiology no one causal approach should drive the questions asked or delimit what counts as useful evidence. Robust causal inference instead comprises a complex narrative, created by scientists appraising, from diverse perspectives, different strands of evidence produced by myriad methods. DAGs can of course be useful, but should not alone wag the causal tale. To make our case, we first address key conceptual issues, after which we offer several concrete examples illustrating how the newly favoured methods, despite their strengths, can also: (i) limit who and what may be deemed a 'cause', thereby narrowing the scope of the field; and (ii) lead to erroneous causal inference, especially if key biological and social assump-

IJEA

International Journal of Epidemiology, 2016, 1–11
doi: 10.1093/ije/dyv341
Original article

Original article

Causality and causal inference in epidemiology: the need for a pluralistic approach

Jan P Vandenbroucke,[1,*,†] Alex Broadbent[2,†] and Neil Pearce[3]

[1]Department of Clinical Epidemiology, Leiden University Medical Center, Leiden, The Netherlands and Department of Clinical Epidemiology, Aarhus University Hospital, Aarhus, Denmark, [2]Department of Philosophy, University of Johannesburg, Auckland Park, South Africa and [3]Department of Medical Statistics and Centre for Global NCDs, London School of Hygiene and Tropical Medicine, London, UK and Centre for Public Health Research, Massey University, Wellington, New Zealand

*Corresponding author. Leiden University Medical Center, Department of Clinical Epidemiology, PO Box 9600, 2300 RC Leiden, The Netherlands. E-mail: J.P.Vandenbroucke@lumc.nl
[†]This paper originated from two separate papers with J.P.V. and A.B. as first authors; they are joint first authors of this paper.

Accepted 25 November 2015

Abstract

Causal inference based on a restricted version of the potential outcomes approach reasoning is assuming an increasingly prominent place in the teaching and practice of epidemiology. The proposed concepts and methods are useful for particular problems, but it would be of concern if the theory and practice of the complete field of epidemiology were to become restricted to this single approach to causal inference. Our concerns are that this theory restricts the questions that epidemiologists may ask and the study designs that they may consider. It also restricts the evidence that may be considered acceptable to assess causality,

UCSF

Current controversies around DAGs

- Restricted Potential Outcomes Approach (RPOA)
  - Causal claims are well-defined only when interventions are well-specified
  - Epidemiologists should only well-defined interventions in order to generate causal hypotheses
  - To investigate causal hypotheses, observational studies should emulate all aspects of experimental studies
- Arguments against RPOA
  - Some interventions are infeasible (e.g. race, sex)
  - Epidemiologists need to study ‘states’, not just interventions
  - Focus on individual study design,
- Need for a pluralistic approach, which includes triangulation, negative controls, and interlocking evidence

UCSF

Lessons for future epidemiologists

1. Causal inference remains a judgment based on integration of diverse types or evidence
2. Diverse strategies to assess causality by ruling out alternatives, such as triangulation, negative controls and interlocking evidence from other types of science
3. Elements of all types of epidemiological study designs, inclusive of those types of design that do not match the ideal counterfactual situation
4. Reflect critically on whether potential biases matter, e.g. for follow-up studies of prevalently exposed persons, or in setting up case-control studies in dynamic populations

Putting it all together

- Causal DAGs is one approach to identify approaches to study design and analysis to eliminate systematic bias
- Confounding is a bias of common causes
  - Causal DAGs allow us to integrate expert knowledge into an analysis to identify true confounders
- Selection bias is a bias of common effects
  - Causal DAGS allow us to identify the structure of selection bias and methods to “adjust” for bias