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

Cindy Leung, ScD, MPH
Center for Health and Community
February 5, 2016

Agenda*

**Material derived from Causal Inference chapters and lecture materials from Dr. Miguel Hernan*

Revisiting Causal Effects

- Neo takes the red pill
  - Five minutes later, he dies
- Had Neo not taken the red pill (and all other things being equal)
  - Five minutes later, he would have been alive
- *Did the red pill have a causal effect?*

Revisiting Causal Effects

- Trinity takes the blue pill
  - Five minutes later, she was alive
- Had Trinity not taken the blue pill (and all other things being equal)
  - Five minutes later, she would have been alive
- *Did the blue pill have a causal effect?*

Revisiting Causal Effects

- We compare…
  - The outcome when the exposure is present
  - The outcome when the exposure is absent
  - Assuming all other conditions are equal

- If these two outcomes differ, we say that the exposure had a causal effect on the outcome

Notation

- **For observed/actual data**:
  - Y=1 if participant died
  - Y=0 if participant was alive
  - A=1 if participant took the red pill
  - A=0 if participant did not take the red pill

ID	A	$Y^{a=0}$	$Y^{a=1}$
Neo	1	0	1
Trinity	0	0	0

- **For ideal/counterfactual data:**
  - $Y^{a=0}$=1 if participant had not taken the red pill, he would have died
  - $Y^{a=1}$=1 if participant had taken the red pill, he would have died

Counterfactual Outcomes

The potential outcomes that would have been observed under each possible treatment value

Or to put it another way…

The outcomes that would have been observed under a treatment value that the subject did not experience

Individual Causal Effects

ID	A	$Y^{a=0}$	$Y^{a=1}$
Neo	1	0	1
Trinity	0	0	0

- Exposure had a causal effect if their outcome, had they been exposed, differs from their outcome, had they been unexposed
  - $Y_i^{a=1} \neq Y_i^{a=0}$
- Generally cannot be determined

Exceptions…

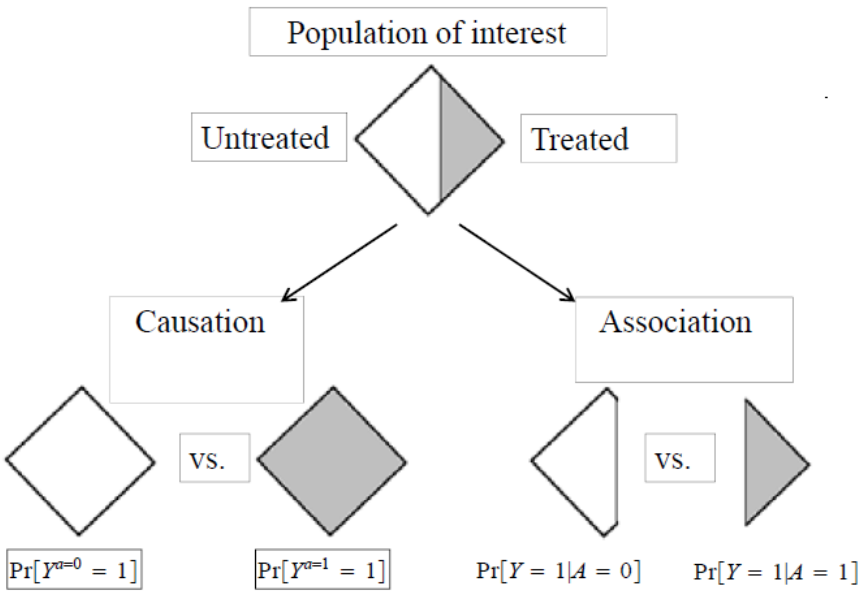
Average Causal Effects

- Exposure had a causal effect if the average risk of the outcome had everyone been exposed, differs from the average risk of the outcome, had no everyone been unexposed
  - $Pr[Y^{a=1}] \neq Pr[Y^{a=0}]$
- Can be determined in ideal randomized studies and in observational studies under strong assumptions

"Association is not causation"

- **Association**: different risk in two disjoint subsets of the population determined by the subjects' actual exposure value
  - Pr[Y=1|A=1] is the risk of Y among those who were exposed (A=1)
  - Pr[Y=1|A=0] is the risk of Y among those who were unexposed (A=0)

- **Causation**: different risk in the entire population under two exposure values
  - Pr[$Y^{a=1}$=1] is the risk of Y had the entire study population been exposed (a=1)
  - Pr[$Y^{a=0}$=1] is the risk of Y had the entire study population been unexposed (a=0)



ORIGINAL CONTRIBUTION

Fish and Omega-3 Fatty Acid Intake and Risk of Coronary Heart Disease in Women

Frank B. Hu, MD
Leslie Bronner, MD
Walter C. Willett, MD
Meir J. Stampfer, MD
Kathryn M. Rexrode, MD
Christine M. Albert, MD
David Hunter, MD
JoAnn E. Manson, MD

Context Higher consumption of fish and omega-3 fatty acids has been associated with a lower risk of coronary heart disease (CHD) in men, but limited data are available regarding women.

Objective To examine the association between fish and long-chain omega-3 fatty acid consumption and risk of CHD in women.

Design, Setting, and Participants Dietary consumption and follow-up data from 84 688 female nurses enrolled in the Nurses' Health Study, aged 34 to 59 years and free from cardiovascular disease and cancer at baseline in 1980, were compared from validated questionnaires completed in 1980, 1984, 1986, 1990, and 1994.

Main Outcome Measures Incident nonfatal myocardial infarction and CHD deaths.

Results During 16 years of follow-up, there were 1513 incident cases of CHD (484 CHD deaths and 1029 nonfatal myocardial infarctions). Compared with women who rarely ate fish (<1 per month), those with a higher intake of fish had a lower risk of CHD. After adjustment for age, smoking, and other cardiovascular risk factors, the multivariable relative risks (RRs) of CHD were 0.79 (95% confidence interval [CI], 0.64-0.97) for fish consumption 1 to 3 times per month, 0.71 (95% CI, 0.58-0.87) for once per week, 0.69 (95% CI, 0.55-0.88) for 2 to 4 times per week, and 0.66 (95% CI, 0.50-0.89) for 5 or more times per week (*P* for trend = .001). Similarly, women with a higher intake of omega-3 fatty acids had a lower risk of CHD, with multivariable RRs of 1.0, 0.93, 0.78, 0.68, and 0.67 (*P*<.001 for trend) across quintiles of intake. For fish intake and omega-3 fatty acids, the inverse association appeared to be stronger for CHD deaths (multivariate RR for fish consumption 5 times per week, 0.55 [95% CI, 0.33-0.90] for CHD deaths vs 0.73 [0.51-1.04]) than for nonfatal myocardial infarction.

Conclusion Among women, higher consumption of fish and omega-3 fatty acids is associated with a lower risk of CHD, particularly CHD deaths.

JAMA. 2002;287:1815-1821 www.jama.com

LOW RATES OF CARDIOVASCULAR disease in populations with a high intake of fish, such as Alaskan Natives,[1,2] Greenland Eskimos,[3,4] and Japanese people residing in fishing villages,[5,6] have suggested that fish consumption may protect against atherosclerosis. Several[7-9] but not all prospective cohort studies[10,11] have found an inverse association between fish consumption and risk of coronary heart disease (CHD). In addition, 2 secondary-prevention trials[12,13] showed that increasing fish consumption or fish-oil supplementation reduced coronary mortality among patients with preexisting coronary disease. Virtually all previous studies on fish consumption and CHD were conducted in men; these results may not apply to women. We therefore

years, follow-up questionnaires were sent to update information and identify new major illnesses. A total of 98 462 women returned the 1980 dietary questionnaire. We excluded

nary revascularization, stroke, or other cardiovascular diseases before 1980. After these exclusions, 84 688 women remained for these analyses.

American Journal of Epidemiology

Vol. 178, No. 3
DOI: 10.1093/aje/kws478
Advance Access publication: June 27, 2013

Original Contribution

Changes in Fish Consumption in Midlife and the Risk of Coronary Heart Disease in Men and Women

Martin Lajous*, Walter C. Willett, James Robins, Jessica G. Young, Eric Rimm, Dariush Mozaffarian, and Miguel A. Hernán

* Correspondence to Dr. Martin Lajous, Kresge Room 820, Department of Epidemiology, Harvard University, 677 Huntington Avenue, Boston, MA 02115 (e-mail: mlajous@post.harvard.edu).

Initially submitted July 5, 2012; accepted for publication December 3, 2012.

Without data from randomized trials, the long-term effects of fish consumption on coronary heart disease (CHD) need to be inferred from observational studies. We estimated CHD risk under different hypothetical interventions on fish consumption during mid- and later life in 2 prospective US cohorts of 25,797 men in the Health Professionals Follow-Up Study and 53,772 women in the Nurses' Health Study. Participants provided information on risk factors and disease every 2 years and on diet every 4 years. We adjusted for baseline and time-varying risk factors for CHD by using the parametric g-formula (where g stands for "generalized"). We observed 1,865 incident CHD cases among men (in 1990–2008) and 1,891 CHD cases among women (in 1986–2008). The risk ratios for CHD when comparing the risk if everyone had consumed at least 2 servings of fish per week with the risk if no one consumed fish during the follow-up periods were 1.03 (95% confidence interval: 0.90, 1.15) for men and 0.87 (95% confidence interval: 0.76, 0.98) for women. Our results suggest that increasing fish consumption to at least 2 servings per week in mid- or later life may lower CHD risk in women but not in men. Our analytical approach allowed us to explicitly specify hypothetical interventions and to assess the effectiveness of dietary changes in midlife.

causal inference; causal model; cohort; coronary artery disease; diet; intervention; lifestyle

Assumptions for causal inference

- Exchangeability
- Positivity
- Consistency (well-defined intervention)
- No model misspecification

Exchangeability

- The risk of the outcome in the exposed would have been the same as the risk of the outcome in the unexposed, had they been exposed
  - And vice versa
- The counterfactual outcome is independent of the exposure

Positivity

- The probability of being exposed must be positive (>0)
  - $Pr[A=a|L=l] > 0$ when $Pr[L=l] > 0$

Consistency

- For each subject, their observed outcome is consistent with one of their counterfactual outcomes

- More specifically, when the exposure is well-defined, the counterfactual outcome is well-defined.
  - When the exposure is not well-defined, the counterfactual outcome is dependent on the exposure manipulation.

Consistency – some examples

- Aspirin and heart attack 
- High sugar diet and diabetes 
- Gender and job discrimination 
- Smoking and lung cancer 
- High blood pressure and mortality 
- Social status and general health 

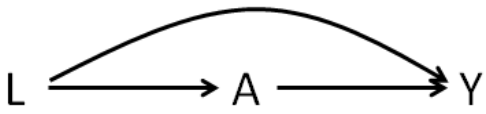
In practice…

	RCTs
Exchangeability	Expected to hold if randomization "works" and large sample size
Positivity	Expected to hold
Well-defined intervention	Expected to hold

Agenda

Causal DAGs

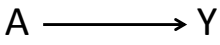
- D: Directed
- A: Acyclic
- G: Graphs



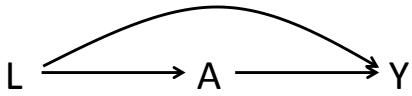
- DAGs are used to:
  - Think conceptually about causal inference
  - Identify sources of bias in a study
  - Identify approaches to study design and analysis to eliminate systematic bias and to help identify average causal effects

Examples of causal DAGs

- Marginally randomized experiment

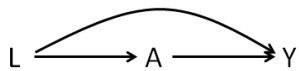


- Conditionally randomized experiment



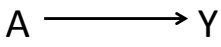
Drawing causal DAGs

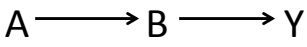
- Identify all random variables of interest/ common causes of exposure and outcome
  - A DAG is causal only if all common causes are shown
  - Variables not common causes do not need to be shown
- Variables drawn in temporal order
- Draw arrows in direction of causation
- Complete DAGs do not make any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



Two variables are associated if…

1) Direct causal effect





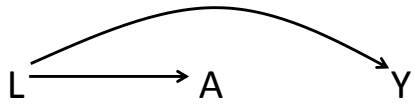
– A: high red meat intake

– B: heme iron

Y: colorectal cancer

Two variables are associated if…

2) Shared common causes



– A: A1 sauce

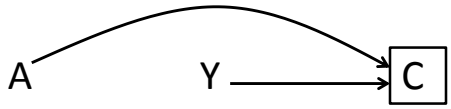
– L: high red meat intake

Y: colorectal cancer

Two variables are associated if…

3) Conditioning on a common effect

– Sometimes called "collider bias"



– A: raining Y: sprinkler turned on

– C: wet grass

Two variables are associated if…

3.1) Conditioning on the descendent of a common effect

– A: raining Y: sprinkler turned on

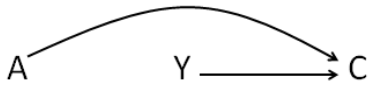
– C: wet grass L: slipped on the grass

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 (regardless of arrow direction)
  - Paths can be blocked or open according to a specific set of rules

d-separation rules

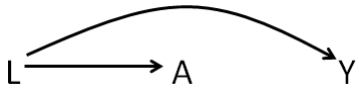
- A path is blocked:
  - If there is a collider



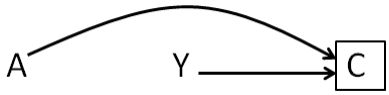
  - When conditioning on a non-collider

d-separation rules

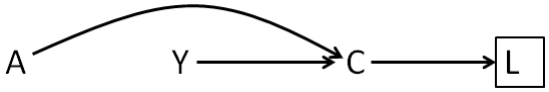
- A path is open:
  - When no non-colliders have been conditioned



  - When conditioning on a collider



  - When conditioning on a descendent of a collider



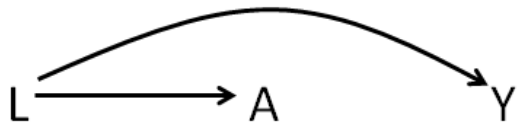
Agenda

Confounders

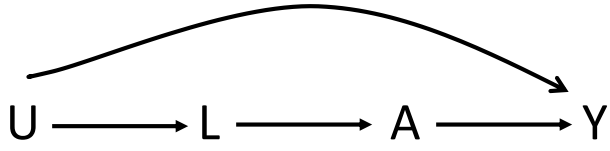
- Standard definition
  - L is associated with A
  - L is associated with Y independent of A
  - L is not on the causal pathway between A and Y
- Collapsibility definition
  - L is a confounder if conditional on L, the association between A and Y is different from the unconditional association
- Causal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

Confounding

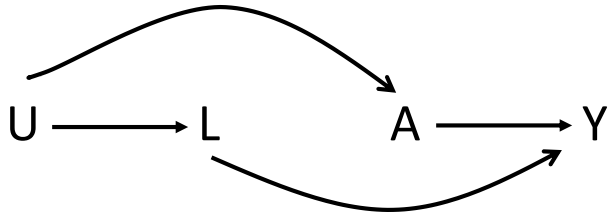
- Bias due to common causes
  - The exposed and unexposed are not exchangeable



Confounder?



Confounder?



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

Control of confounding

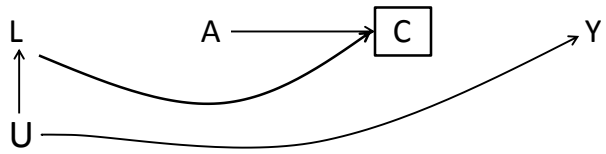
- In the design phase:
  - Randomization
  - Restriction
  - Matching
- In the analytic phase:
  - Stratification
  - IPW

Agenda

Selection bias

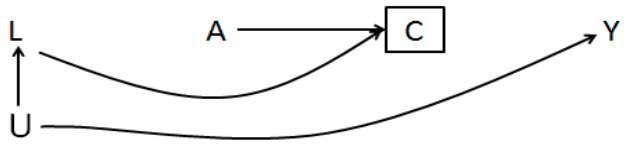
- Systematic bias due to the selection of subjects into the study or the analysis
  - Conditioning on a common effect
    - Survival
    - Diagnosis with disease
    - Selection into a study
    - Complete data

Differential loss to Follow-Up



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

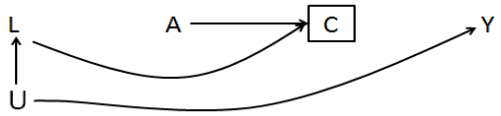
Volunteer/ self-selection



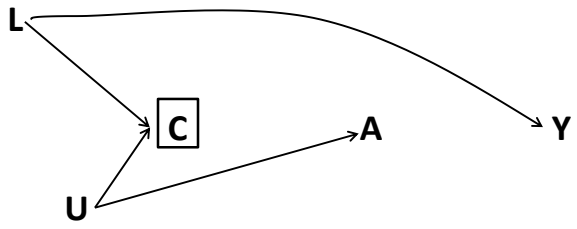
- A: Smoking
- Y: Heart disease
- C: Volunteer to participate
- L: Awareness of heart disease
- U: Family history of heart disease

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

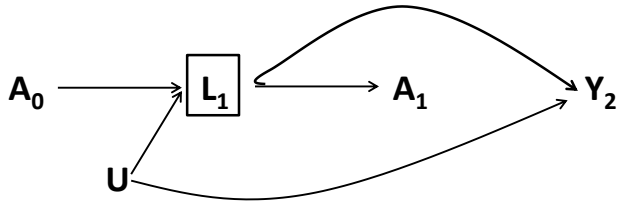


Confounding or Selection Bias?



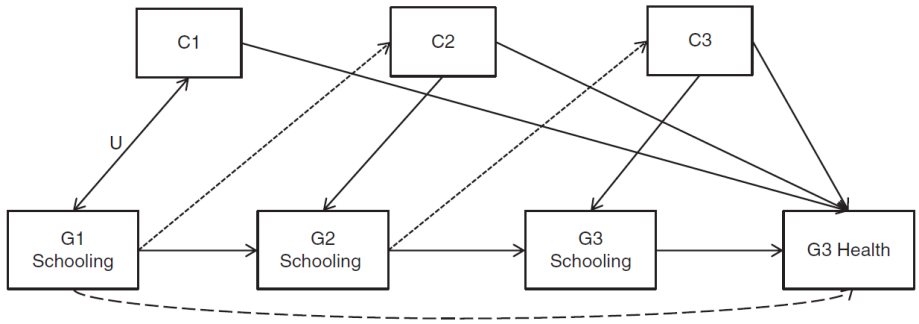
- A: Physical Activity
- Y: Heart disease
- C: Being a firefighter
- L: Parental SES
- U: Attraction to activities that involve physical activity

Confounding or Selection Bias?



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

Example from HW



Importance of DAGs

- Two distinct causal structures that lead to systematic bias:
  - Confounding (common causes)
  - Selection bias (conditioning on common effects)
- Drawing DAGs help to provide a clear picture of the various biases that may be present in a study

Example

You are part of a team evaluating the effects of a **mindful eating intervention** on **excessive gestational weight gain**. Women were not randomized though 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?

Putting it all together

- Estimation of average causal effects requires strong assumptions
- Causal DAGs can be used 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