

harvard

## CAUSES AND COUNTERFACTUALS

Judea Pearl  
University of California  
Los Angeles  
(www.cs.ucla.edu/~judea/jsm09)

atlanta

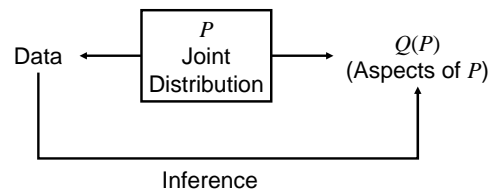
## THEORETICAL DEVELOPMENTS IN CAUSAL INFERENCE

Judea Pearl  
University of California  
Los Angeles  
(www.cs.ucla.edu/~judea/jsm09)

### OUTLINE

- Inference: Statistical vs. Causal, distinctions, and mental barriers
- Unified conceptualization of counterfactuals, structural-equations, and graphs
- Inference to three types of claims:
  1. Effect of potential interventions
  2. Attribution (Causes of Effects)
  3. Direct and indirect effects
- Frills

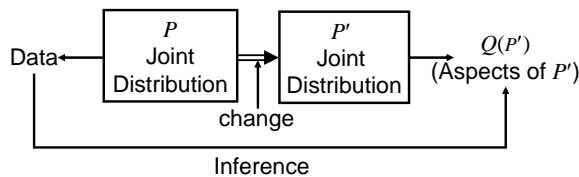
### TRADITIONAL STATISTICAL INFERENCE PARADIGM



e.g.,  
Infer whether customers who bought product A would also buy product B.  
 $Q = P(B | A)$

### FROM STATISTICAL TO CAUSAL ANALYSIS: 1. THE DIFFERENCES

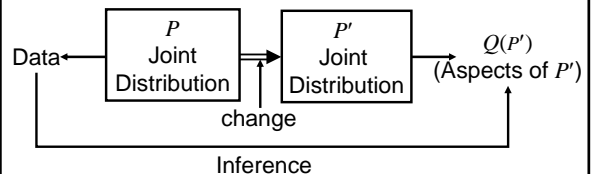
Probability and statistics deal with static relations



What happens when  $P$  changes?  
e.g.,  
Infer whether customers who bought product A would still buy A if we were to double the price.

### FROM STATISTICAL TO CAUSAL ANALYSIS: 1. THE DIFFERENCES

What remains invariant when  $P$  changes say, to satisfy  $P'(price=2)=1$



Note:  $P'(v) \neq P(v | price = 2)$   
 $P$  does not tell us how it ought to change  
e.g. Curing symptoms vs. curing diseases  
e.g. Analogy: mechanical deformation

**FROM STATISTICAL TO CAUSAL ANALYSIS:  
1. THE DIFFERENCES (CONT)**

- Causal and statistical concepts do not mix.
 

<b>CAUSAL</b>	<b>STATISTICAL</b>
Spurious correlation	Regression
Randomization / Intervention	Association / Independence
Confounding / Effect	"Controlling for" / Conditioning
Instrumental variable	Odd and risk ratios
Strong Exogeneity	Collapsibility / Granger causality
Explanatory variables	Propensity score
- 
- 
- 

**FROM STATISTICAL TO CAUSAL ANALYSIS:  
2. MENTAL BARRIERS**

- Causal and statistical concepts do not mix.
 

<b>CAUSAL</b>	<b>STATISTICAL</b>
Spurious correlation	Regression
Randomization / Intervention	Association / Independence
Confounding / Effect	"Controlling for" / Conditioning
Instrumental variable	Odd and risk ratios
Strong Exogeneity	Collapsibility / Granger causality
Explanatory variables	Propensity score
- No causes in – no causes out (Cartwright, 1989)
 

statistical assumptions + data } ⇒ causal conclusions  
causal assumptions
- Causal assumptions cannot be expressed in the mathematical language of standard statistics.
- 

**FROM STATISTICAL TO CAUSAL ANALYSIS:  
2. MENTAL BARRIERS**

- Causal and statistical concepts do not mix.
 

<b>CAUSAL</b>	<b>STATISTICAL</b>
Spurious correlation	Regression
Randomization / Intervention	Association / Independence
Confounding / Effect	"Controlling for" / Conditioning
Instrumental variable	Odd and risk ratios
Strong Exogeneity	Collapsibility / Granger causality
Explanatory variables	Propensity score
- No causes in – no causes out (Cartwright, 1989)
 

statistical assumptions + data } ⇒ causal conclusions  
causal assumptions
- Causal assumptions cannot be expressed in the mathematical language of standard statistics.
- Non-standard mathematics:
  - Structural equation models (Wright, 1920; Simon, 1960)
  - Counterfactuals (Neyman-Rubin ( $Y_x$ ), Lewis ( $x \rightarrow Y$ ))

**WHY CAUSALITY NEEDS  
SPECIAL MATHEMATICS**

Scientific Equations (e.g., Hooke's Law) are non-algebraic  
e.g., Length (Y) equals a constant (2) times the weight (X)  
Correct notation:

<del><math>Y = 2X</math></del>	$X = 1$
$X = 1$	$Y = 2$
<u>Process information</u>	<u>The solution</u>

Had X been 3, Y would be 6.  
If we raise X to 3, Y would be 6.  
Must "wipe out"  $X = 1$ .

**WHY CAUSALITY NEEDS  
SPECIAL MATHEMATICS**

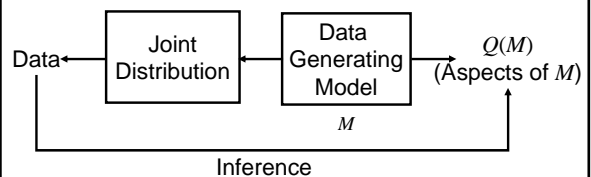
Scientific Equations (e.g., Hooke's Law) are non-algebraic  
e.g., Length (Y) equals a constant (2) times the weight (X)  
Correct notation:

(or)

$Y \leftarrow 2X$	$X = 1$
$X = 1$	$Y = 2$
<u>Process information</u>	<u>The solution</u>

Had X been 3, Y would be 6.  
If we raise X to 3, Y would be 6.  
Must "wipe out"  $X = 1$ .

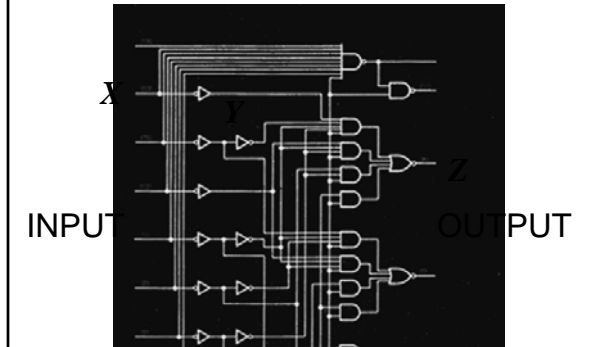
**THE STRUCTURAL MODEL  
PARADIGM**



$M$  – Invariant strategy (mechanism, recipe, law, protocol) by which Nature assigns values to variables in the analysis.

“Think Nature, not experiment!”

## FAMILIAR CAUSAL MODEL ORACLE FOR MANIPULATION



## STRUCTURAL CAUSAL MODELS

Definition: A structural causal model is a 4-tuple  $\langle V, U, F, P(u) \rangle$ , where

- $V = \{V_1, \dots, V_n\}$  are endogeneous variables
- $U = \{U_1, \dots, U_m\}$  are background variables
- $F = \{f_1, \dots, f_n\}$  are functions determining  $V$ ,  
 $v_i = f_i(v, u)$  e.g.,  $y = \alpha + \beta x + u_Y$
- $P(u)$  is a distribution over  $U$

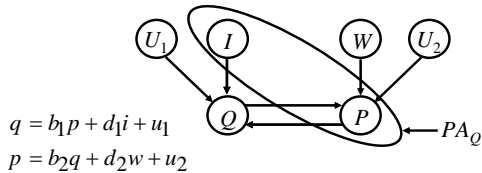
$P(u)$  and  $F$  induce a distribution  $P(v)$  over observable variables

## STRUCTURAL MODELS AND CAUSAL DIAGRAMS

The functions  $v_i = f_i(v, u)$  define a graph

$$v_i = f_i(p, a, u_i) \quad PA_i \subseteq V \setminus V_i \quad U_i \subseteq U$$

Example: Price – Quantity equations in economics

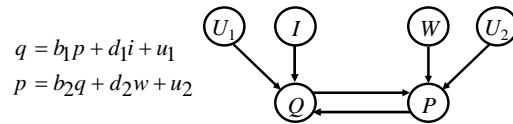


## STRUCTURAL MODELS AND INTERVENTION

Let  $X$  be a set of variables in  $V$ .

The action  $do(x)$  sets  $X$  to constants  $x$  regardless of the factors which previously determined  $X$ .

$do(x)$  replaces all functions  $f_i$  determining  $X$  with the constant functions  $X=x$ , to create a mutilated model  $M_x$ .

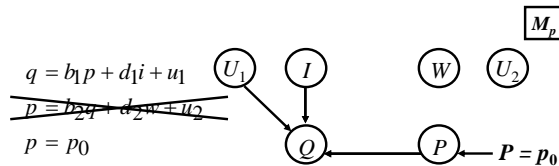


## STRUCTURAL MODELS AND INTERVENTION

Let  $X$  be a set of variables in  $V$ .

The action  $do(x)$  sets  $X$  to constants  $x$  regardless of the factors which previously determined  $X$ .

$do(x)$  replaces all functions  $f_i$  determining  $X$  with the constant functions  $X=x$ , to create a mutilated model  $M_x$ .



## CAUSAL MODELS AND COUNTERFACTUALS

Definition:

The sentence: "Y would be y (in situation u), had X been x," denoted  $Y_x(u) = y$ , means:

The solution for Y in a mutilated model  $M_x$ , (i.e., the equations for X replaced by  $X = x$ ) with input  $U = u$ , is equal to y.

The Fundamental Equation of Counterfactuals:

$$Y_x(u) = Y_{M_x}(u)$$

## CAUSAL MODELS AND COUNTERFACTUALS

Definition:

The sentence: "Y would be y (in situation u), had X been x," denoted  $Y_x(u) = y$ , means:

The solution for Y in a mutilated model  $M_x$  (i.e., the equations for X replaced by  $X = x$ ) with input  $U = u$ , is equal to y.

- Joint probabilities of counterfactuals:

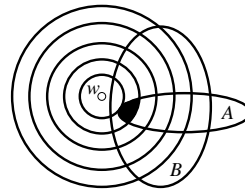
$$P(Y_x = y, Z_w = z) = \sum_{u: Y_x(u)=y, Z_w(u)=z} P(u)$$

In particular:

$$P(y | do(x)) \stackrel{\Delta}{=} P(Y_x = y) = \sum_{u: Y_x(u)=y} P(u)$$

$$PN(Y_{x'} = y' | x, y) = \sum_{u: Y_{x'}(u)=y'} P(u | x, y)$$

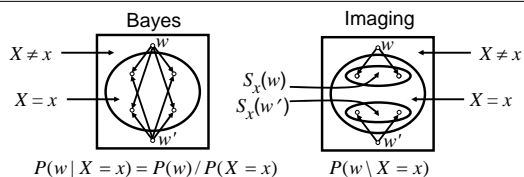
## STRUCTURAL AND SIMILARITY-BASED COUNTERFACTUALS



Lewis's account (1973): The counterfactual  $A \square \rightarrow B$  (read: "B if it were A") is true in a world w just in case B is true in all the closest A-worlds to w.

Structural account (1995): The counterfactual  $Y_x(u) = y$  (read: "Y=y if X were x") is true in situation u just in case  $Y_{M_x}(u) = y$ .

## BAYESIAN AND IMAGING CONDITIONALIZATIONS



The  $do(x)$  operator is an imaging operator provided:  
 Provision 1: Worlds with equal histories should be considered equally similar to any given world.  
 Provision 2: Equally-similar worlds should receive mass in proportion to their prior probabilities (Bayesian tie-breaking)

## AXIOMS OF STRUCTURAL COUNTERFACTUALS

$Y_x(u) = y$ : Y would be y, had X been x (in state  $U = u$ ) (Galles, Pearl, Halpern, 1998):

- Definiteness  
 $\exists x \in X \text{ s.t. } X_y(u) = x$
- Uniqueness  
 $(X_y(u) = x) \& (X_{y'}(u) = x') \Rightarrow x = x'$
- Effectiveness  
 $X_{xw}(u) = x$
- Composition (generalized consistency)  
 $X_w(u) = x \Rightarrow Y_{wx}(u) = Y_w(u)$
- Reversibility  
 $(Y_{xw}(u) = y) \& (W_{xy}(u) = w) \Rightarrow Y_x(u) = y$

## REGRESSION VS. STRUCTURAL EQUATIONS (THE CONFUSION OF THE CENTURY)

Regression (claimless, nonfalsifiable):

$$Y = ax + \varepsilon_y$$

Structural (empirical, falsifiable):

$$Y = bx + u_y$$

Claim: (regardless of distributions):

$$E(Y | do(x)) = E(Y | do(x), do(z)) = bx$$

The mothers of all questions:

Q. When would b equal a?

A. When all back-door paths are blocked,  $(u_y \perp\!\!\!\perp X)$

Q. When is b estimable by regression methods?

A. Graphical criteria available

## THE FOUR NECESSARY STEPS OF CAUSAL ANALYSIS

Define: Express the target quantity Q as a function  $Q(M)$  that can be computed from any model M.

Assume: Formulate causal assumptions using ordinary scientific language and represent their structural part in graphical form.

Identify: Determine if Q is identifiable.

Estimate: Estimate Q if it is identifiable; approximate it, if it is not.



## FORMULATING ASSUMPTIONS THREE LANGUAGES

1. English: Smoking ( $X$ ), Cancer ( $Y$ ), Tar ( $Z$ ), Genotypes ( $U$ )

2. Counterfactuals:  $Z_x(u) = Z_{yx}(u)$ ,

$$X_y(u) = X_{zy}(u) = X_z(u) = X(u),$$

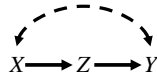
$$Y_z(u) = Y_{zx}(u),$$

$$Z_x \perp\!\!\!\perp \{Y_z, X\}$$

Not too friendly:

Consistent?, complete?, redundant?, arguable?

3. Structural:



## IDENTIFIABILITY

Definition:

Let  $Q(M)$  be any quantity defined on a causal model  $M$ , and let  $A$  be a set of assumption.

$Q$  is identifiable relative to  $A$  iff

$$P(M_1) = P(M_2) \Rightarrow Q(M_1) = Q(M_2)$$

for all  $M_1, M_2$ , that satisfy  $A$ .

- 
- 

## IDENTIFIABILITY

Definition:

Let  $Q(M)$  be any quantity defined on a causal model  $M$ , and let  $A$  be a set of assumption.

$Q$  is identifiable relative to  $A$  iff

$$P(M_1) = P(M_2) \Rightarrow Q(M_1) = Q(M_2)$$

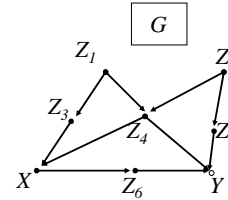
for all  $M_1, M_2$ , that satisfy  $A$ .

In other words,  $Q$  can be determined uniquely from the probability distribution  $P(v)$  of the endogenous variables,  $V$ , and assumptions  $A$ .

$A$  is displayed in graph  $G$ .

## THE PROBLEM OF CONFOUNDING

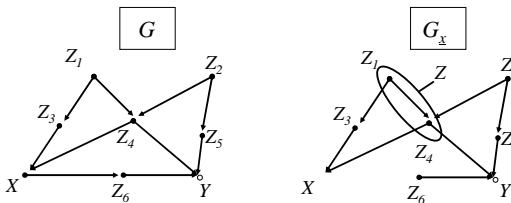
Find the effect of  $X$  on  $Y$ ,  $P(y|do(x))$ , given the causal assumptions shown in  $G$ , where  $Z_1, \dots, Z_k$  are auxiliary variables.



Can  $P(y|do(x))$  be estimated if only a subset,  $Z$ , can be measured?

## ELIMINATING CONFOUNDING BIAS THE BACK-DOOR CRITERION

$P(y | do(x))$  is estimable if there is a set  $Z$  of variables such that  $Z$   $d$ -separates  $X$  from  $Y$  in  $G_x$ .

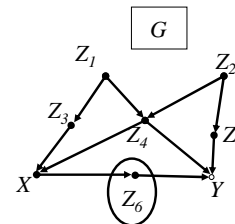


Moreover,  $P(y | do(x)) = \sum_z P(y | x, z) P(z)$   
("adjusting" for  $Z$ )

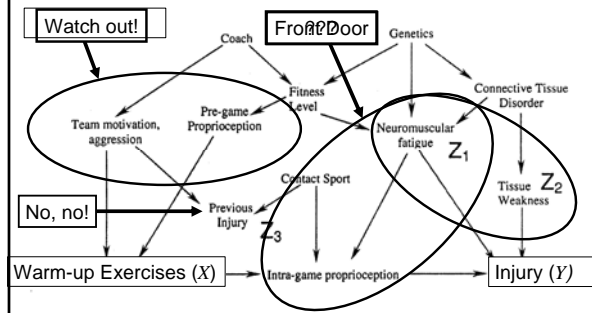
## EFFECT OF INTERVENTION BEYOND ADJUSTMENT

Theorem (Tian-Pearl 2002)

We can identify  $P(y|do(x))$  if there is no child  $Z$  of  $X$  connected to  $X$  by a confounding path.



## EFFECT OF WARM-UP ON INJURY (After Shrier & Platt, 2008)



## EFFECT OF INTERVENTION COMPLETE IDENTIFICATION

- Complete calculus for reducing  $P(y|do(x), z)$  to expressions void of *do*-operators.
- Complete graphical criterion for identifying causal effects (Shpitser and Pearl, 2006).
- Complete graphical criterion for empirical testability of counterfactuals (Shpitser and Pearl, 2007).

## COUNTERFACTUALS AT WORK ETT – EFFECT OF TREATMENT ON THE TREATED

1. Regret:  
I took a pill to fall asleep.  
Perhaps I should not have?
2. Program evaluation:  
What would terminating a program do to those enrolled?

## THE FOUR NECESSARY STEPS EFFECT OF TREATMENT ON THE TREATED

Define: Express the target quantity  $Q$  as a function  $Q(M)$  that can be computed from any model  $M$ .

$$ETT \triangleq P(Y_x = y | X = x')$$

Assume: Formulate causal assumptions using ordinary scientific language and represent their structural part in graphical form.

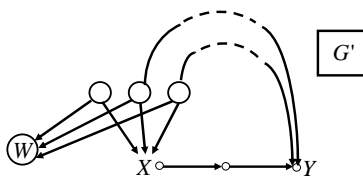
Identify: Determine if  $Q$  is identifiable.

Estimate: Estimate  $Q$  if it is identifiable; approximate it, if it is not.

## ETT - IDENTIFICATION

Theorem (Shpitser-Pearl, 2009)

ETT is identifiable in  $G$  iff  $P(y | do(x), w)$  is identifiable in  $G'$

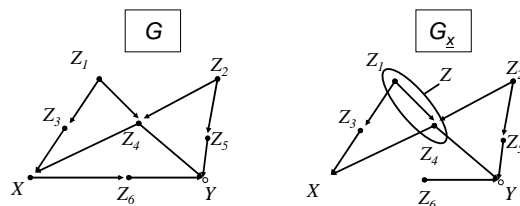


Moreover,  $ETT = P(Y_x = y | x') = P(y | do(x), w)$

Complete graphical criterion

## ETT - THE BACK-DOOR CRITERION

$P(Y_x = y | x')$  is identifiable in  $G$  if there is a set  $Z$  of variables such that  $Z$  *d*-separates  $X$  from  $Y$  in  $G_x$ .



Moreover,  $ETT = \sum_z P(y | x, z) P(z | x')$

◦ "Standardized morbidity"

## FROM IDENTIFICATION TO ESTIMATION

- Define:** Express the target quantity  $Q$  as a function  $Q(M)$  that can be computed from any model  $M$ .
- $$Q = P(y | do(x))$$
- Assume:** Formulate causal assumptions using ordinary scientific language and represent their structural part in graphical form.
- Identify:** Determine if  $Q$  is identifiable.
- Estimate:** Estimate  $Q$  if it is identifiable; approximate it, if it is not.

## PROPENSITY SCORE ESTIMATOR (Rosenbaum & Rubin, 1983)

$P(y | do(x)) = ?$

$$L(z_1, z_2, z_3, z_4, z_5) \triangleq P(X = 1 | z_1, z_2, z_3, z_4, z_5)$$

Theorem:  $\sum_z P(y | z, x)P(z) = \sum_l P(y | L = l, x)P(L = l)$

Adjustment for  $L$  replaces Adjustment for  $Z$

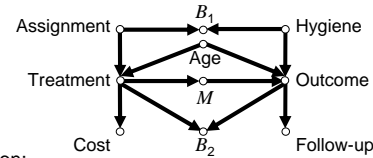
## WHAT PROPENSITY SCORE (PS) PRACTITIONERS NEED TO KNOW

$$L(z) = P(X = 1 | Z = z)$$

$$\sum_z P(y | z, x)P(z) = \sum_l P(y | l, x)P(l)$$

1. The asymptotic bias of PS is EQUAL to that of ordinary adjustment (for same  $Z$ ).
2. Including an additional covariate in the analysis CAN SPOIL the bias-reduction potential of others.
3. Choosing sufficient set for PS, requires knowledge about the model.

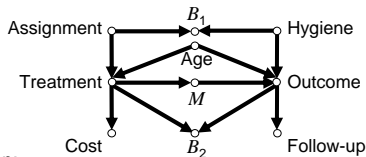
## WHICH COVARIATES MAY / SHOULD BE ADJUSTED FOR?



**Question:** Which of these eight covariates may be included in the propensity score function (for matching) and which should be excluded.

**Answer:**  
 Must include: Age  
 Must exclude:  $B_1, M, B_2, \text{Follow-up, Assignment without Age}$   
 May include: Cost, Hygiene, {Assignment + Age}, {Hygiene + Age +  $B_1$ }, more . . .

## WHICH COVARIATES MAY / SHOULD BE ADJUSTED FOR?



**Question:** Which of these eight covariates may be included in the propensity score function (for matching) and which should be excluded.

**Answer:**  
 Must include: Age  
 Must exclude:  $B_1, M, B_2, \text{Follow-up, Assignment without Age}$   
 May include: Cost, Hygiene, {Assignment + Age}, {Hygiene + Age +  $B_1$ }, more . . .

## WHAT PROPENSITY SCORE (PS) PRACTITIONERS NEED TO KNOW

$$L(z) = P(X = 1 | Z = z)$$

$$\sum_z P(y | z, x)P(z) = \sum_l P(y | l, x)P(l)$$

1. The asymptotic bias of PS is EQUAL to that of ordinary adjustment (for same  $Z$ ).
2. Including an additional covariate in the analysis CAN SPOIL the bias-reduction potential of others.
3. Choosing sufficient set for PS, requires knowledge about the model.
4. That any empirical test of the bias-reduction potential of PS, can only be generalized to cases where the causal relationships among covariates, observed and unobserved is the same.

## TWO PARADIGMS FOR CAUSAL INFERENCE

Observed:  $P(X, Y, Z, \dots)$   
 Conclusions needed:  $P(Y_x = y), P(X_y = x | Z = z) \dots$

How do we connect observables,  $X, Y, Z, \dots$  to counterfactuals  $Y_x, X_z, Z_y, \dots$  ?

### N-R model

Counterfactuals are primitives, new variables

Super-distribution

$P^*(X, Y, \dots, Y_x, X_z, \dots)$

$X, Y, Z$  constrain  $Y_x, Z_y, \dots$

### Structural model

Counterfactuals are derived quantities

Subscripts modify the model and distribution

$P(Y_x = y) = P_{M_x}(Y = y)$

## "SUPER" DISTRIBUTION IN N-R MODEL

X	Y	Z	$Y_{x=0}$	$Y_{x=1}$	$X_{z=0}$	$X_{z=1}$	$X_{y=0} \dots$	U
0	0	0	0	1	0	0	0...	$u_1$
0	1	1	1	0	1	0	1...	$u_2$
0	0	0	1	0	0	1	1...	$u_3$
1	0	0	0	0	0	1	0...	$u_4$

inconsistency:  $x=0 \Rightarrow Y_{x=0} = Y$       $Y = xY_1 + (1-x)Y_0$


Defines:  $P^*(X, Y, Z, \dots, Y_x, Z_y, \dots, Y_{xz}, Z_{xy}, \dots)$   
 $P^*(Y_x = y | Z, X_z)$   
 $Y_x \perp\!\!\!\perp X | Z_y$

## THE FOUR NECESSARY STEPS IN POTENTIAL-OUTCOME FRAMEWORK

- Define: Express the target quantity  $Q$  as a counterfactual formula
- Assume: Formulate causal assumptions using the distribution:  
 $P(X | Y, Z, Y(1), Y(0))$
- Identify: Determine if  $Q$  is identifiable.
- Estimate: Estimate  $Q$  if it is identifiable; approximate it, if it is not.

## GRAPHICAL – COUNTERFACTUALS SYMBIOSIS

Every causal graph expresses counterfactual assumptions, e.g.,  $X \rightarrow Y \rightarrow Z$

1. Missing arrows  $Y \leftarrow Z$
2. Missing arcs 

consistent, and readable from the graph.

Every theorem in SCM is a theorem in Potential-Outcome Model, and conversely.

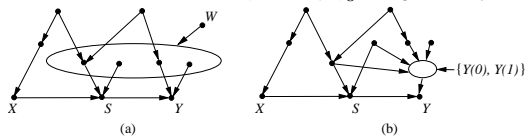
## DEMISTIFYING STRONG IGNORABILITY

$$\{Y(0), Y(1)\} \perp\!\!\!\perp X | Z \quad (\text{Ignorability})$$

$$P(y | do(x)) = \sum_z P(y | z, x) P(z) \quad (\text{Z-admissibility})$$

$$(X \perp\!\!\!\perp Y | Z)_{G_X} \quad (\text{Back-door})$$

Is there a  $W$  in  $G$  such that  $(W \perp\!\!\!\perp X | Z)_G \Rightarrow$  Ignorability?



## DETERMINING THE CAUSES OF EFFECTS (The Attribution Problem)

- Your Honor! My client (Mr. A) died BECAUSE he used that drug.



## DETERMINING THE CAUSES OF EFFECTS (The Attribution Problem)

- Your Honor! My client (Mr. A) died BECAUSE he used that drug.



- Court to decide if it is MORE PROBABLE THAN NOT that A would be alive BUT FOR the drug!  
 $PN = P(? | A \text{ is dead, took the drug}) \geq 0.50$

## THE ATTRIBUTION PROBLEM

Definition:

- What is the meaning of  $PN(x,y)$ :  
"Probability that event  $y$  would not have occurred if it were not for event  $x$ , given that  $x$  and  $y$  did in fact occur."

Answer:

$$PN(x, y) = P(Y_{x'} = y' | x, y)$$

Computable from  $M$

## THE ATTRIBUTION PROBLEM

Definition:

- What is the meaning of  $PN(x,y)$ :  
"Probability that event  $y$  would not have occurred if it were not for event  $x$ , given that  $x$  and  $y$  did in fact occur."

Identification:

- Under what condition can  $PN(x,y)$  be learned from statistical data, i.e., observational, experimental and combined.

## TYPICAL THEOREMS

(Tian and Pearl, 2000)

- Bounds given combined nonexperimental and experimental data

$$\max \left\{ \frac{0}{P(x,y)} \right\} \leq PN \leq \min \left\{ \frac{1}{P(x,y)} \right\}$$

- Identifiability under monotonicity (Combined data)

$$PN = \frac{P(y/x) - P(y/x')}{P(y/x)} + \frac{P(y/x') - P(y/x')}{P(x,y)}$$

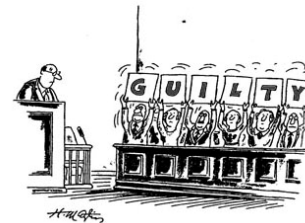
corrected Excess-Risk-Ratio

## CAN FREQUENCY DATA DECIDE LEGAL RESPONSIBILITY?

	Experimental		Nonexperimental	
	$do(x)$	$do(x')$	$x$	$x'$
Deaths ( $y$ )	16	14	2	28
Survivals ( $y'$ )	984	986	998	972
	1,000	1,000	1,000	1,000

- Nonexperimental data: drug usage predicts longer life
- Experimental data: drug has negligible effect on survival
- Plaintiff: Mr. A is special.
  - He actually died
  - He used the drug by choice
- Court to decide (given both data):  
Is it more probable than not that A would be alive but for the drug?

## SOLUTION TO THE ATTRIBUTION PROBLEM



- WITH PROBABILITY ONE  $1 \leq P(y'_{x'} | x, y) \leq 1$
- Combined data tell more than each study alone

## EFFECT DECOMPOSITION (direct vs. indirect effects)

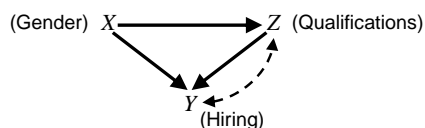
1. Why decompose effects?
2. What is the definition of direct and indirect effects?
3. What are the policy implications of direct and indirect effects?
4. When can direct and indirect effect be estimated consistently from experimental and nonexperimental data?

## WHY DECOMPOSE EFFECTS?

1. To understand how Nature works
2. To comply with legal requirements
3. To predict the effects of new type of interventions:  
Signal routing, rather than variable fixing

## LEGAL IMPLICATIONS OF DIRECT EFFECT

Can data prove an employer guilty of hiring discrimination?



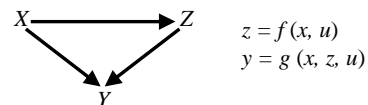
What is the direct effect of  $X$  on  $Y$ ?

$$E(Y | do(x_1), do(z)) - E(Y | do(x_0), do(z))$$

(averaged over  $z$ ) Adjust for  $Z$ ? No! No!

## NATURAL INTERPRETATION OF AVERAGE DIRECT EFFECTS

Robins and Greenland (1992) – “Pure”



$$z = f(x, u)$$

$$y = g(x, z, u)$$

Natural Direct Effect of  $X$  on  $Y$ :  $DE(x_0, x_1; Y)$

The expected change in  $Y$ , when we change  $X$  from  $x_0$  to  $x_1$ , and, for each  $u$ , we keep  $Z$  constant at whatever value it attained before the change.

$$E[Y_{x_1 Z_{x_0}} - Y_{x_0}]$$

In linear models,  $DE =$  Controlled Direct Effect

## DEFINITION AND IDENTIFICATION OF NESTED COUNTERFACTUALS

Consider the quantity  $Q \triangleq E_u[Y_{XZ_{X^*}}(u)]$

Given  $\langle M, P(u) \rangle$ ,  $Q$  is well defined

Given  $u$ ,  $Z_{x^*}(u)$  is the solution for  $Z$  in  $M_{x^*}$ , call it  $z$

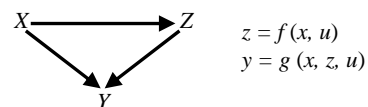
$Y_{XZ_{x^*}}(u)$  is the solution for  $Y$  in  $M_{xz}$

Can  $Q$  be estimated from  $\left\{ \begin{array}{l} \text{experimental} \\ \text{nonexperimental} \end{array} \right\}$  data?

Experimental: nest-free expression

Nonexperimental: subscript-free expression

## DEFINITION OF INDIRECT EFFECTS



$$z = f(x, u)$$

$$y = g(x, z, u)$$

Indirect Effect of  $X$  on  $Y$ :  $IE(x_0, x_1; Y)$

The expected change in  $Y$  when we keep  $X$  constant, say at  $x_0$ , and let  $Z$  change to whatever value it would have attained had  $X$  changed to  $x_1$ .

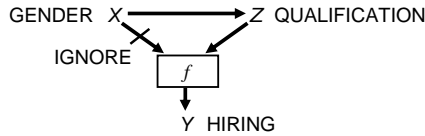
$$E[Y_{x_0 Z_{x_1}} - Y_{x_0}]$$

In linear models,  $IE = TE - DE$

## POLICY IMPLICATIONS OF INDIRECT EFFECTS

What is the indirect effect of X on Y?

The effect of Gender on Hiring if sex discrimination is eliminated.



Blocking a link – a new type of intervention

## EXPERIMENTAL IDENTIFICATION OF NATURAL DIRECT EFFECTS

Theorem: If there exists a set  $W$  such that

$$Y_{xz} \perp\!\!\!\perp Z_{x^*} \mid W \text{ for all } z \text{ and } x$$

Then the average direct effect

$$DE(x, x^*; Y) = E(Y_{x, Z_{x^*}}) - E(Y_{x^*})$$

Is identifiable from experimental data and is given by

$$DE(x, x^*; Y) = \sum_{w, z} [E(Y_{xz} \mid w) - E(Y_{x^*z} \mid w)] P(Z_{x^*} = z \mid w) P(w)$$

## GRAPHICAL CONDITION FOR EXPERIMENTAL IDENTIFICATION OF DIRECT EFFECTS

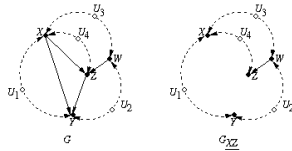
Theorem: If there exists a set  $W$  such that

$$(Y \perp\!\!\!\perp Z \mid W)_{G_{XZ}} \text{ and } W \subseteq ND(X \cup Z)$$

then,

$$DE(x, x^*; Y) = \sum_{w, z} [E(Y_{xz} \mid w) - E(Y_{x^*z} \mid w)] P(Z_{x^*} = z \mid w) P(w)$$

Example:



## MEDIATION FORMULAS

1. The natural direct and indirect effects are identifiable in Markovian models, and are given by:

$$DE = \sum_z [E(Y \mid do(x', z)) - E(Y \mid do(x, z))] P(z \mid do(x))$$

$$IE = \sum_z E(Y \mid do(x, z)) [P(z \mid do(x')) - P(z \mid do(x))]$$

3. All *do*-expressions are estimable by regression.

## CONCLUSIONS

It is **not** as simple as it seems

• on an explicit causal structure that is defensible on scientific grounds.

• Unification of the graphical, potential-outcome and structural equation approaches (Aristotle, 384-322 B.C.)

• Friendly and formal solutions to century-old problems and confusions.

From Charlie Poole

## QUESTIONS???

They will be answered