WHY THERE IS NO STATISTICAL TEST FOR CONFOUNDING, WHY MANY THINK THERE IS, AND WHY THEY ARE ALMOST RIGHT

Judea Pearl
Cognitive Systems Laboratory
Computer Science Department
UCLA

Definition 6.2.1
(No-Confounding; Causal Definition)
Let M be a causal model of the data-generating process—that is, a formal description of how the value of each observed variable is determined. Denote by $P(y|do(x))$ the probability of the response event $Y = y$ under the hypothetical intervention $X = x$, calculated according to M. We say that X and Y are not confounded in M if and only if

$$P(y|do(x)) = P(y|x) \qquad \textbf{(6.10)}$$

for all x and y in their respective domains, where $P(y|x)$ is the conditional probability generated by M.

Definition 6.2.2
(No-Confounding: Associational Criterion)
Let T be the set of variables in a problem that are not affected by X. We say that X and Y are not confounded in the presence of T if every member Z of T satisfies at least one of the following):

(U_1) Z is not associated with X, (i.e., $P(x|z) = P(x)$);

(U_2) Z is not associated with Y conditional on X (i.e., $P(y|z, x) = P(y|x)$).

Example 6.3.1

Imagine a situation where **exposure** (X) is influenced by a person's **education** (E), **disease** (Y) is influenced by both exposure and **age** (A), and **car type** (Z) is influenced by both **age** (A) and **education** (E). These relationships are shown schematically in Figure 6.3

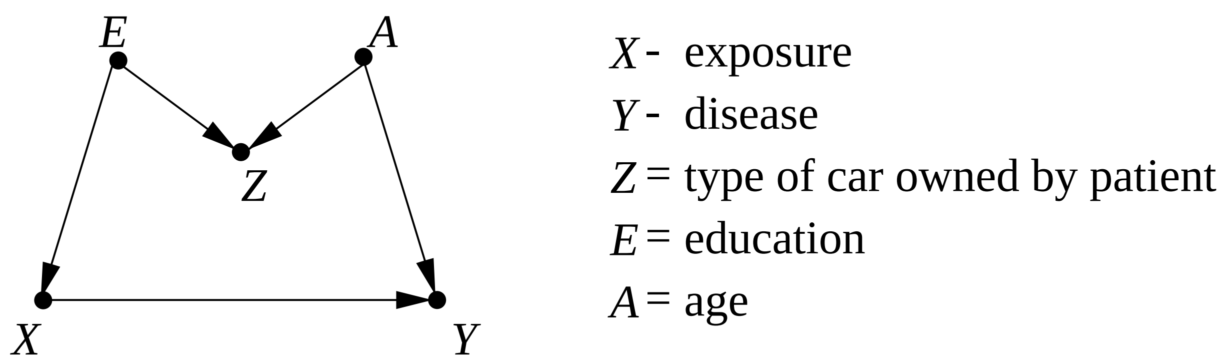

Figure 6.3

Both (U_1) and (U_2) fail on Z, yet X and Y are not confounded.

Moreover, adjusting for Z would yield a **biased** result

$$P(y|do(x)) \neq \sum_z P(y|x,z)P(z)$$

MODIFIED ASSOCIATIONAL CRITERION

Definition 6.3.2
(No-Confounding; Modified Associational Criterion)
Let T be the set of variables in a problem that are not affected by X but **may potentially affect $\boldsymbol{Y}$**. We say that X and Y are unconfounded by the presence of T if and only if every member Z of T satisfies either (U_1) or (U_2) of Definition 6.2.2.

(U_1) Z is not associated with X,
(i.e., $P(x|z) = P(x)$),

(U_2) Z is not associated with Y within strata of X (i.e., $P(y|z,x) = P(y|x)$).

Example 6.3.3

Consider a causal model defined by the linear equations:

$$x = \alpha z + \epsilon_1 \qquad \textbf{(6.11)}$$

$$y = \beta x + \gamma z + \epsilon_2 \qquad \textbf{(6.12)}$$

where ϵ_1 and ϵ_2 are correlated unmeasured variables having $cov(\epsilon_1, \epsilon_2) = r$ and where Z is an exogenous variable that is uncorrelated with ϵ_1 or ϵ_2.

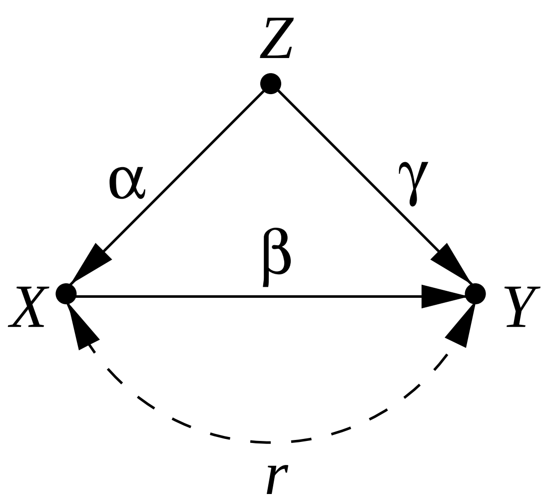

Figure 6.4

If $r = -\alpha\gamma$ then $r_{YX} = \beta$.
Thus, (X, Y) is not confounded, though (U_1) and (U_2) are violated.

STABLE UNBIASEDNESS

Definition 6.4.1
(Stable Unbiasedness)
Let A be a set of assumptions (or restrictions) on the data-generating process, and let C_A be a class of causal models satisfying A. The effect estimate of X on Y is said to be **stably unbiased** given A if $P_M(y|do(x)) = P(y|x)$ holds in every model M in C_A. Correspondingly, we say that the pair (X, Y) is **stably unconfounded**, given A.

EXAMPLE OF MODEL ASSUMPTIONS: GRAPHS

Definition 6.4.2
Structurally Stable No-Confounding)
Let A_D be the set of assumptions embedded in a causal diagram D. We say that X and Y are **stably unconfounded** given A_D if $P(y|do(x)) = P(y|x)$ holds in every parameterization of D.
By "parameterization" we mean an assignment of functions to the links of the diagram and prior probabilities to the background variables in the diagram.

Theorem 6.4.3 (Common-Cause Principle) Let A_D be the set of assumptions embedded in an acyclic causal diagram D. Variables X and Y are stably unconfounded given A_D if and only if X and Y have no common ancestor in D.

(More precisely, All back-door paths from X to Y are blocked.)

Theorem 6.4.4 (Criterion for Stable No-Confounding)
Let A_Z denote the assumptions that (i) the data are generated by some (unspecified) acyclic model M and (ii) Z is a variable in M that is unaffected by X but may possibly affect Y. If both of the associational criteria (U_1) and (U_2) of Definition 6.2.2 are violated, then (X, Y) are not stably unconfounded given A_Z.

By "possibly affecting Y" we mean: A_Z does not contain the assumption that Z does not affect Y. In other words, the diagram associated with M must contain a directed path from Z to Y.

Finding just *any* variable Z that satisfies A_Z and violates (U_1) and (U_2) permits us to disqualify (X, Y) as stably unconfounded (though (X, Y) may be incidentally unconfounded in the particular experimental conditions prevailing in the study).

COLLAPSIBILITY
VS.
NONCONFOUNDING

Definition 6.5.1
(Collapsibility)
Let $g[P(y,x)]$ be any functional that measures the association between Y and X in the joint distribution $P(x,y)$. We say that g is **collapsible** on a variable Z if

$$E_z g[P(x,y|z) = g[P(x,y)]$$

Corollary 6.5.2
(Stable No-Confounding Implies Collapsibility)
Let Z be any variable that is not affected by X and that may possibly affect Y. Let $g[P(x,y)]$ be any linear functional that measures the association between X and Y. If g is not collapsible on Z, then X and Y are not stably unconfounded.