Using Causal Diagrams to Understand Covariate Balance in Propensity Score Models

Onyebuchi A. Arah
Department of Epidemiology
UCLA School of Public Health
Los Angeles, CA
&
Department of Social Medicine
Academic Medical Center, University of Amsterdam
Amsterdam, The Netherlands

Society for Epidemiologic Research (SER)
Anaheim, June 2009

BACKGROUND

The propensity score (PS) is increasingly being used to estimate causal effects in observational studies.

The PS quantifies the probability that a subject receives the exposure or treatment being studied. The PS maps numerous covariates into one score which substitutes for the high-dimensional covariate set in subsequent analysis of the effect of exposure on the outcome.

A central concern in any PS application is achieving covariate balance between exposure groups. Balancing (all) covariates is seen by many as an important yardstick for judging how successful a PS application is.

For high-dimensional covariate set **Z**, treatment *X*, and outcome *Y*, the PS is estimated as $e(\mathbf{Z}) = \Pr(X=1|\mathbf{Z})$.

If **Z** is sufficient to block all biasing (confounding) paths between X and Y, and the PS, $e(\mathbf{Z})$, is successful in rendering *X* and **Z** independent, then $e(\mathbf{Z})$ should be able to block the biasing paths between X and Y.

OBJECTIVE

This study shows how to use traditional causal diagrams as well as those augmented with the functional mapping encoded in the propensity score to explain covariate balance in propensity score models.

METHODS

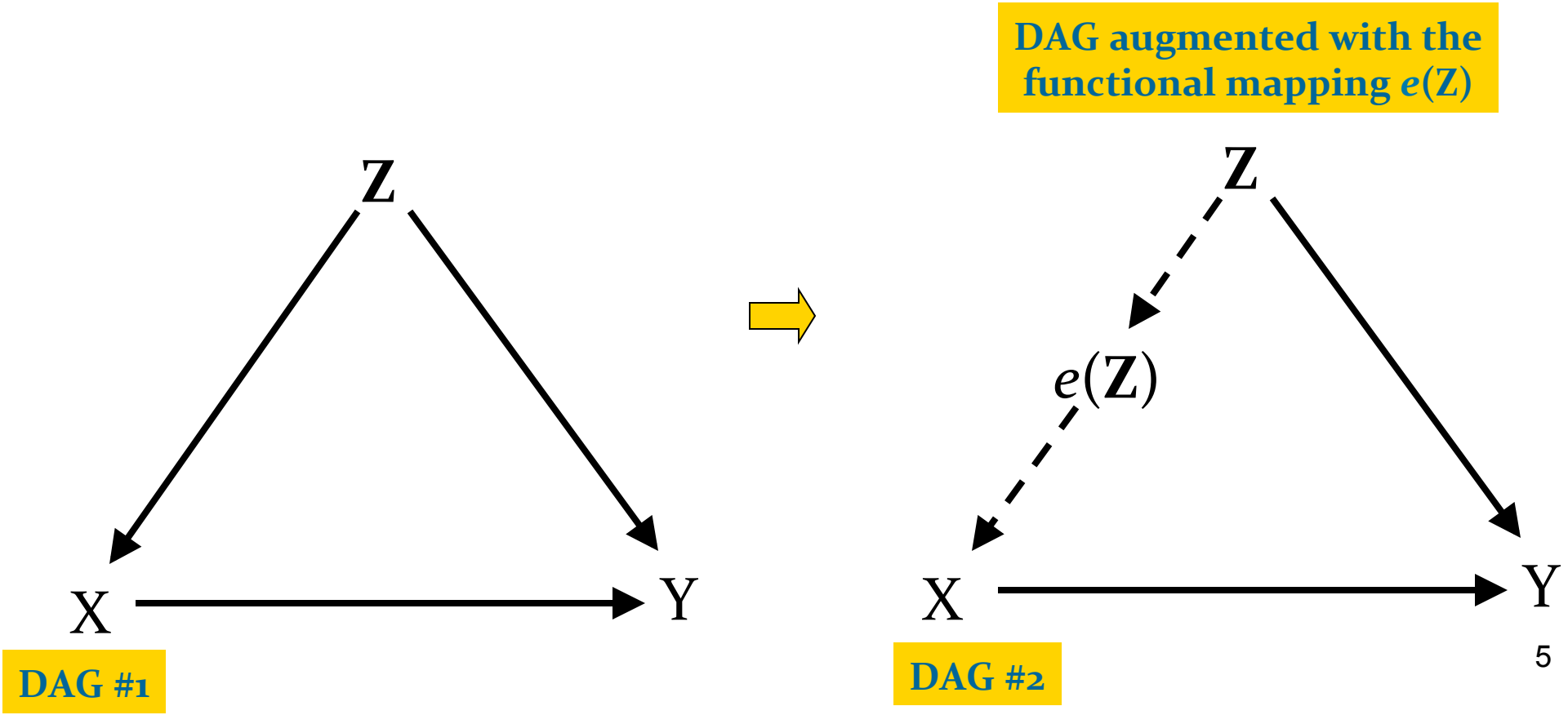
The researcher uses functional mapping augmentation to replace **solid** edges between the exposure and other covariates with **dashed** edges interconnecting the selected covariates (**Z**) with $e(\mathbf{Z})$ and subsequently $e(\mathbf{Z})$ with the exposure (X): **Z**→X becomes **Z**---> $e(\mathbf{Z})$--->X.

The graphical rules of causal diagrams combined with the conditional independence properties of the propensity score can then be applied to check for covariate balance.

Simple Confounding

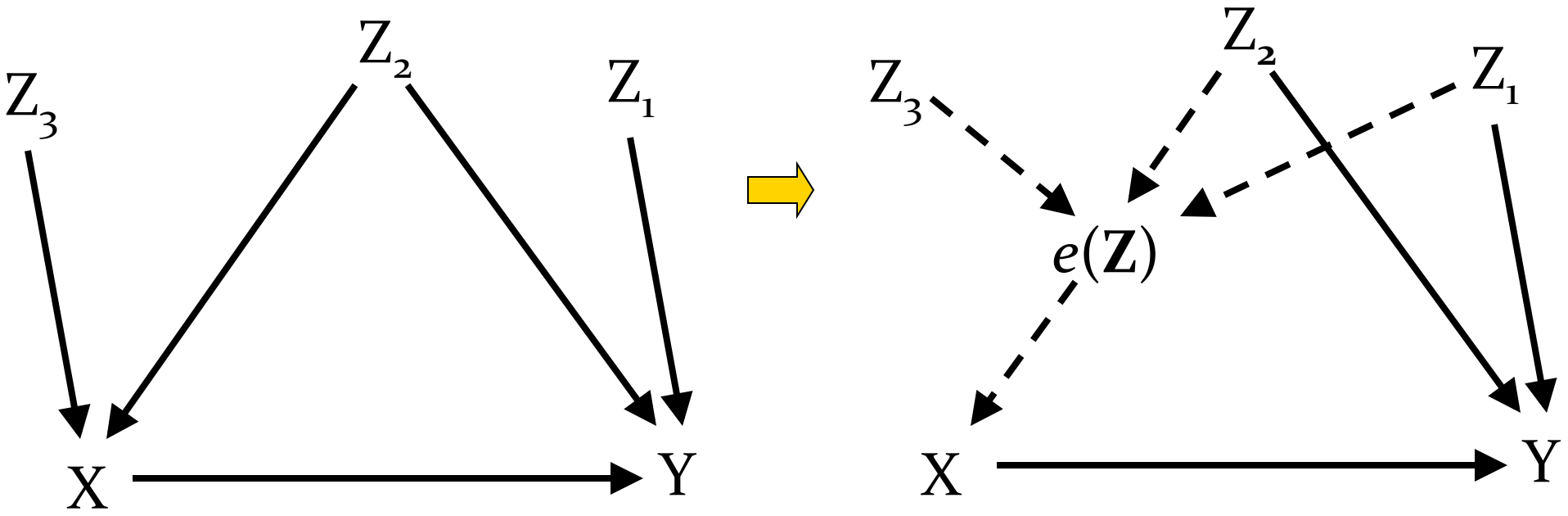
Path X-**Z**-Y sufficiently blocked by Z, and if the edge connecting Z and X is successfully functionally mapped as $e(\mathbf{Z})$, then

- given $e(\mathbf{Z})$, the selected covariate **Z** will be balanced within X, and
- $e(\mathbf{Z})$ should block the path X-$e(\mathbf{Z})$-Z-Y if PS correctly applied.



Confounding variables plus outcome-only and exposure-only predictors

- Prior to applying the PS model (DAG #3), the outcome-only predictor Z_1 set (but not the exposure-only predictor Z_3 or confounding variable Z_2) is balanced within X. Neither Z_1 nor Z_3 is on a biasing path between X and Y.
- When all three covariate types are included in the PS model $e(\mathbf{Z})$, all three should balanced within X given $e(\mathbf{Z})$. See DAG#4.



DAG #3

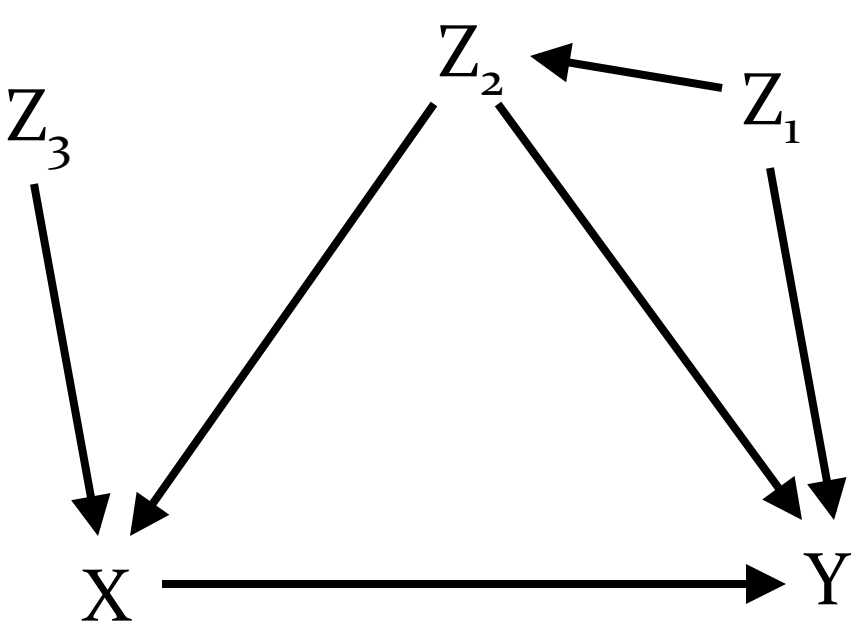
DAG #4

- **Potential covariate balance between exposed and unexposed after applying (e.g. matching on) PS**

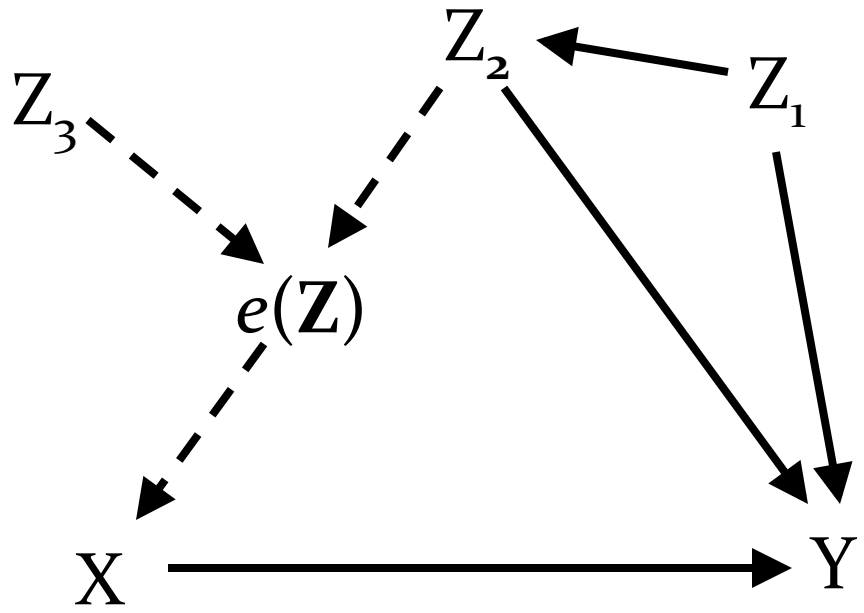
PS model $e(Z)$	Balance of Z_3 (variables related only to treatment)	Balance of Z_2 (confounders, i.e. variables related to both treatment and outcome)	Balance of Z_1 (variables related only to outcome)
$P(X=1\|Z_2=z_2)$	No	Yes	Yes
$P(X=1\|Z_2=z_2, Z_1=z_1)$	No	Yes	Yes
$P(X=1\|Z_2=z_2, Z_3=z_3)$	Yes	Yes	Yes

Confounding variables plus outcome-confounder and exposure-only predictors

- Prior to applying the PS model (DAG #5), the outcome Y and the confounder Z_2 share a common cause Z_1 which, without controlling for Z_2, is not balanced within X. Z_2 can block the biasing path left open by Z_1 and Z_2 between X and Y.
- If Z_2 (but perhaps not Z_1) were included in the PS model $e(\mathbf{Z})$, then Z_1 can be expected to be balanced within X given $e(\mathbf{Z})$. See DAG#6.



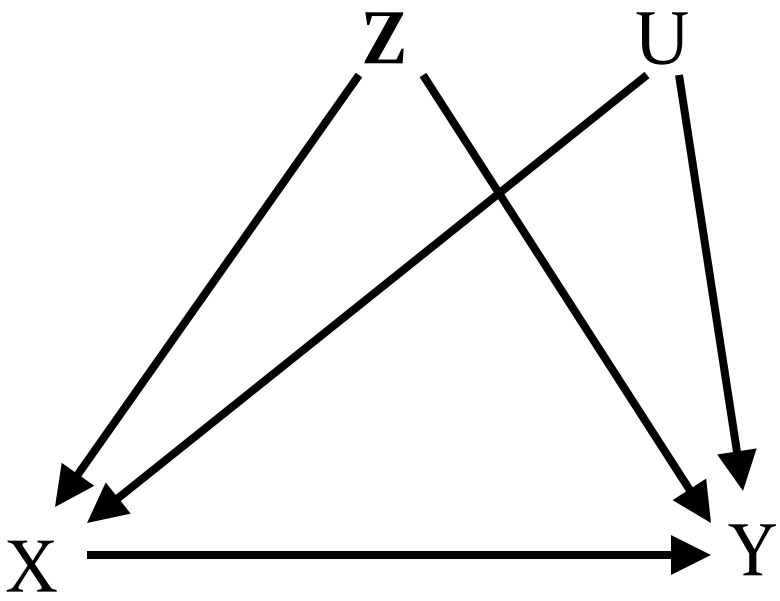
DAG #5



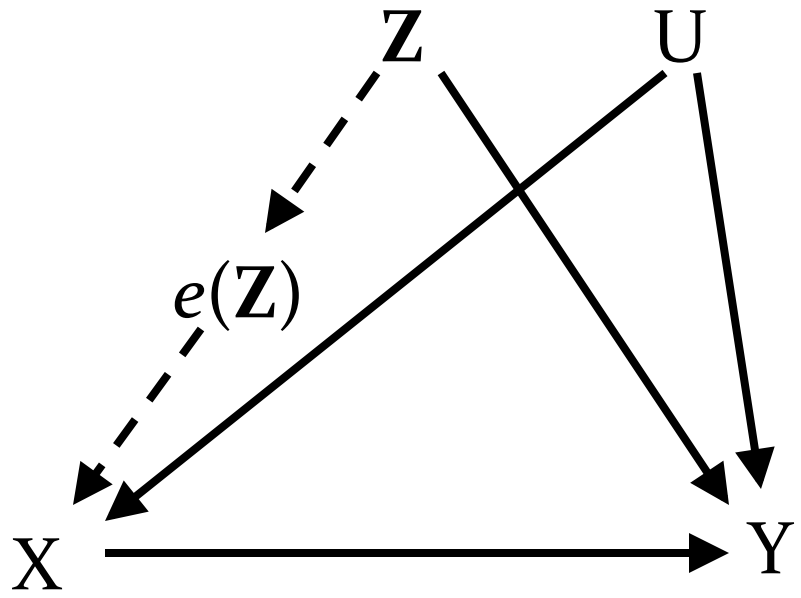
DAG #6

Uncontrolled Confounding

- Representing uncontrolled confounding is also straightforward: when a covariate U is unmeasured (DAG#5) and is, therefore, not included in the PS model $e(\mathbf{Z})$, the biasing or backdoor path left open by the uncontrolled U remains open even after applying $e(\mathbf{Z})$ as would be intuitively expected (DAG #6).



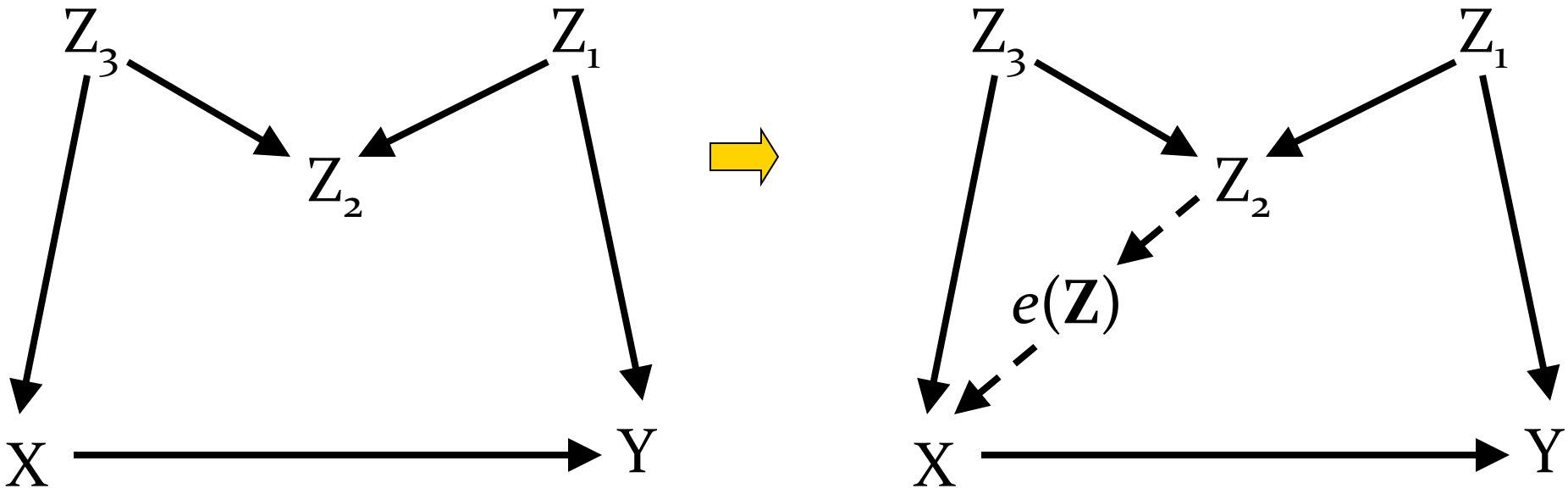
DAG #7



DAG #8

M-Bias Structure

- It is well-known that adjusting for just the collider Z_2 in DAG #9 opens up a biasing path X-Z_1-Z_2-Z_3-Y between X and Y.
- Now, if the collider Z_2 (but not the covariates Z_1 and Z_3) were included in the PS model $e(\mathbf{Z})$, then a similar collider bias would be induced in the relationship between X and Y. This happens because adjusting for $e(\mathbf{Z})$ which acts as a "descendant" of Z_2 opens up the colliding path at Z_2, also leading to marginal dependence between Z_1 and X (DAG #10).

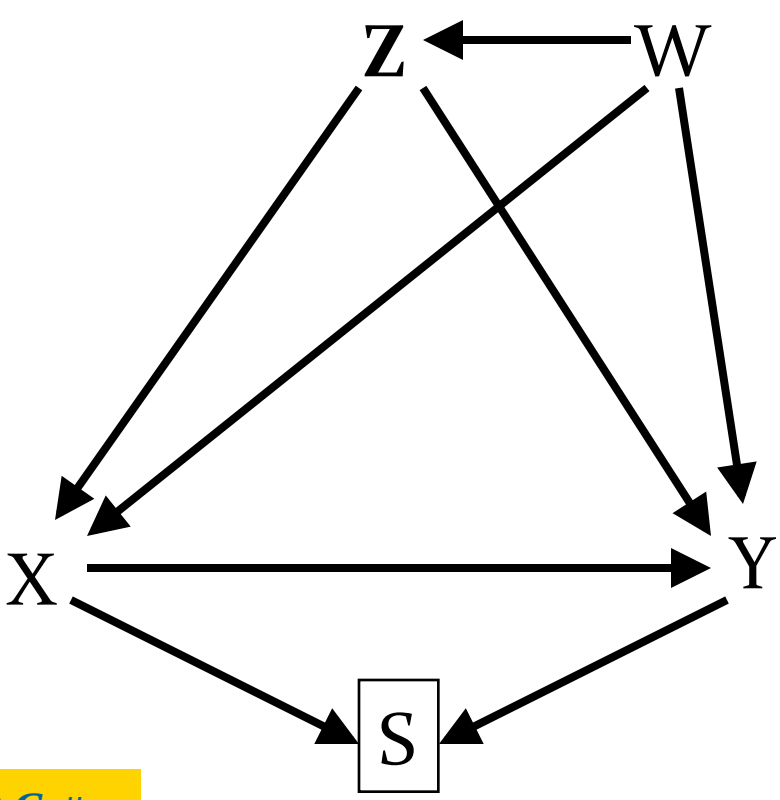


DAG #9

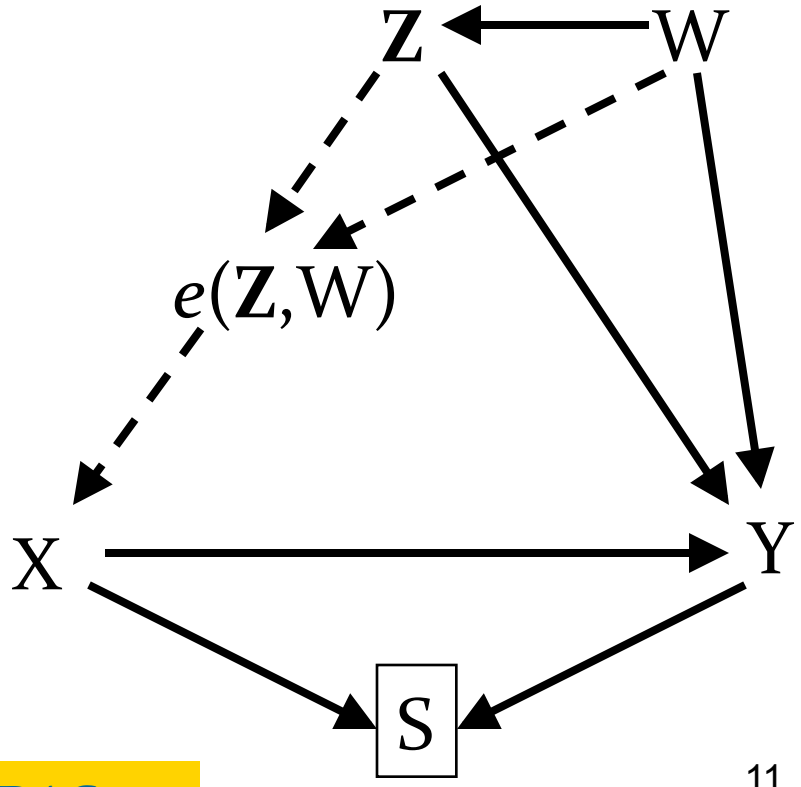
DAG #10

Selection Bias

- If there is selection bias (from, say, loss to follow-up in cohort study) as depicted in DAG#11, that selection bias will still persist even after applying the PS $e(\mathbf{Z},W)$ which adjusts for the confounders **Z** and W (DAG #12).



DAG #11



DAG #12

CONCLUSIONS

Using both traditional and augmented causal diagrams, this study demonstrates how outcome-only predictors are balanced within the exposure whether or not they are included in the propensity score model.

Other types of variables such as confounders and exposure-only predictors must be included in the propensity score model in order to achieve covariate balance between the exposed and unexposed. Nonetheless, only imbalances in covariates which are on confounding paths lead to biased estimates of the causal effect of the exposure. These theoretical results were supported by extensive Monte Carlo simulations.

Funding Acknowledgement:

The Netherlands Organization for Scientific Research (NWO)

Presenter:

Onyebuchi A. Arah

MD, MSc, DSc, MPH, PhD

Department of Epidemiology

University of California, Los Angeles (UCLA)

School of Public Health

Los Angeles, CA 90095-1772

Telephone: 310 206 0050

Fax: 310 206 6039

Email: arah@ucla.edu