Augmented Causal Diagrams for Measurement Error

Onyebuchi A. Arah

Department of Epidemiology
UCLA School of Public Health
Los Angeles, California
&
Department of Public Health
Academic Medical Center, University of Amsterdam
The Netherlands

Society for Epidemiologic Research
Seattle, 26 June 2010

Observational studies often plagued by multiple biases such as from **measurement error (ME)**, selection, and unmeasured confounders.

Yet, these biases are rarely quantitatively assessed or even acknowledged.

ME easily the most difficult of the biases to handle.

Here, this presentation demonstrates how to

1. Depict ME using augmented directed acyclic graphs (DAGs)
2. Visualize the biasing paths induced by ME
3. Visualize ME adjustment or modeling methods

Notation

(V) denotes unmeasured true variable V

ε_V is the disturbance term in V

V* is measured or misclassified V

ε_{V^*} is the disturbance term in V*

Solid arrow: true (unmeasured) association between any two variables

Dashed arrow: measured association (perhaps externally available)

Bidirected dashed arc: common cause present

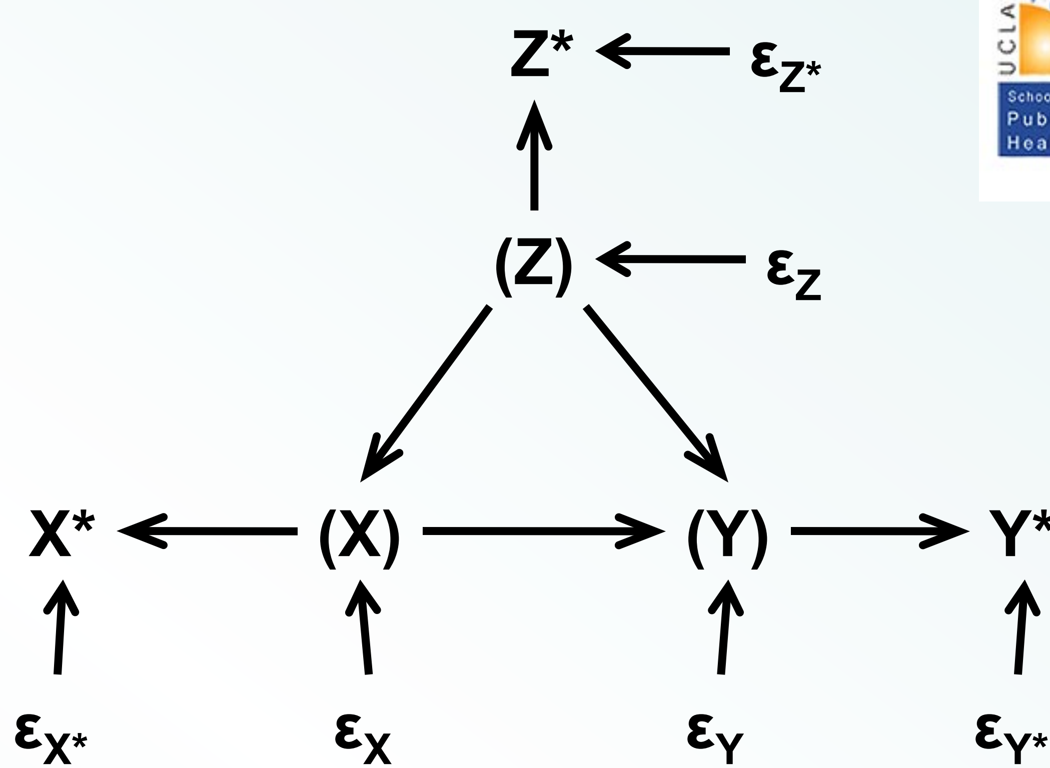
X is exposure, Z confounder, and Y outcome

Depicting and Visualizing Independent Non-differential Measurement Error

Although typically omitted from DAGs, disturbance terms are important in measurement error DAGs.

ME is **independent:** the disturbance terms are (marginally) uncorrelated.

ME is **nondifferential:** the only arrows into V* are from the disturbance term and true unmeasured variable.



Each red dashed arrow is equivalent to all the open paths between the two variables connected by the red arrow.

Z*-X* association is given by:

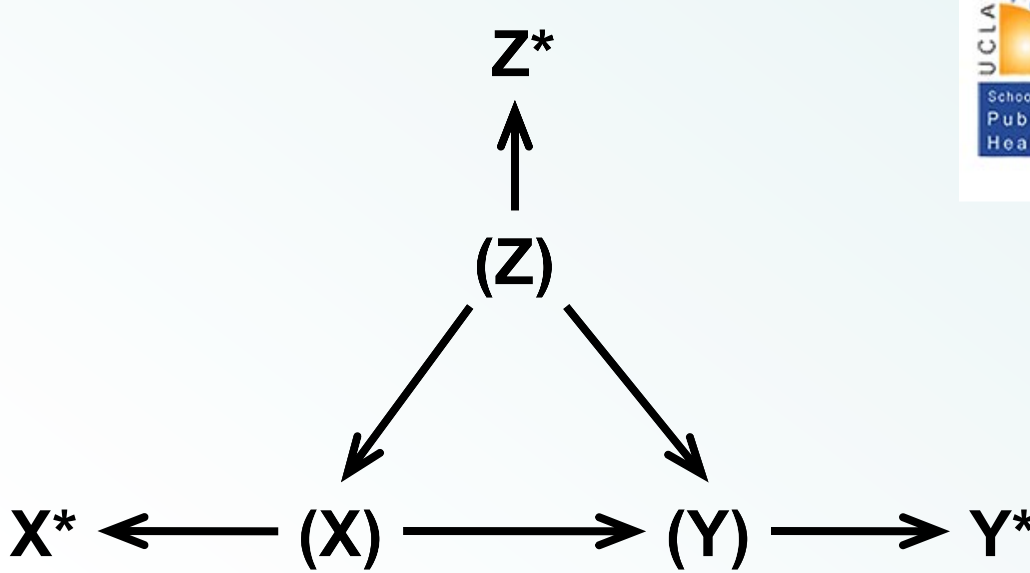
1. Z*-(Z)-(X)-X*.

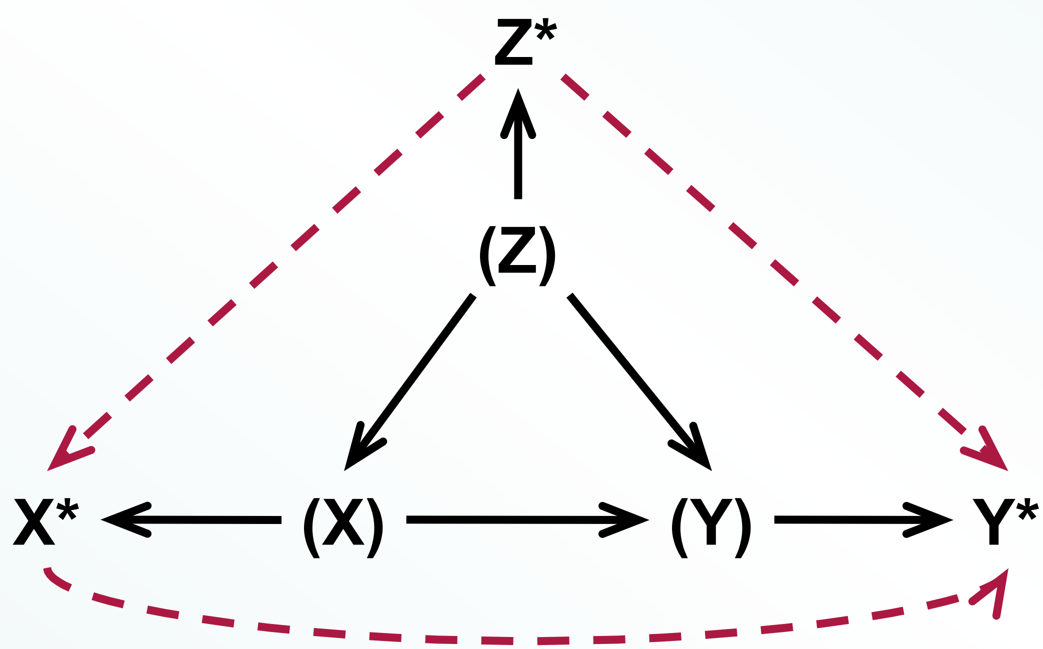
Z*-Y* association by:

1. Z*-(Z)-(Y)-Y*
2. Z*-(Z)-(X)-(Y)-Y*.

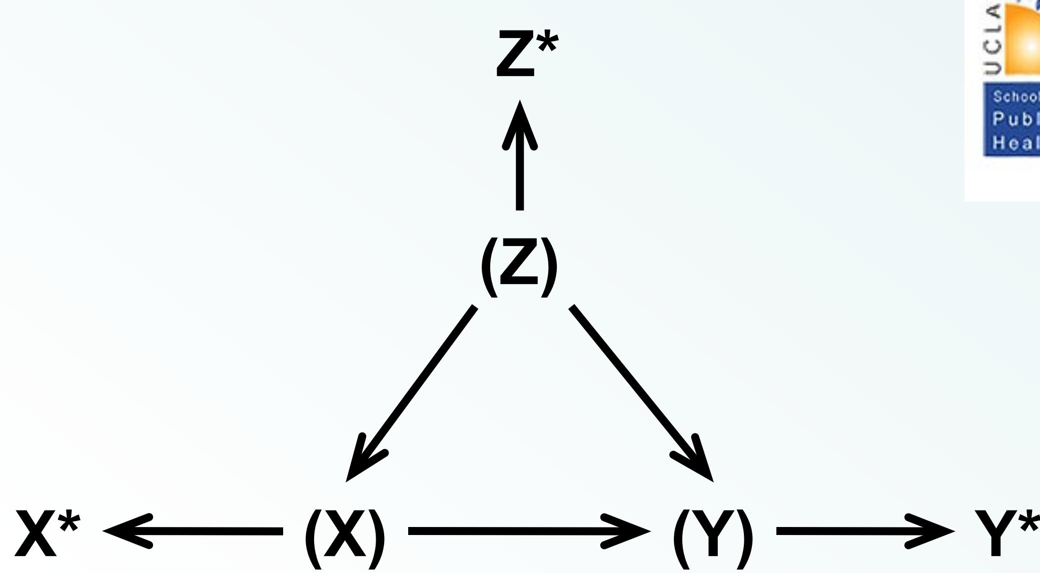
X*-Y* association by:

1. X*-(X)-(Y)-Y*
2. X*-(X)-(Z)-(Y)-Y*.

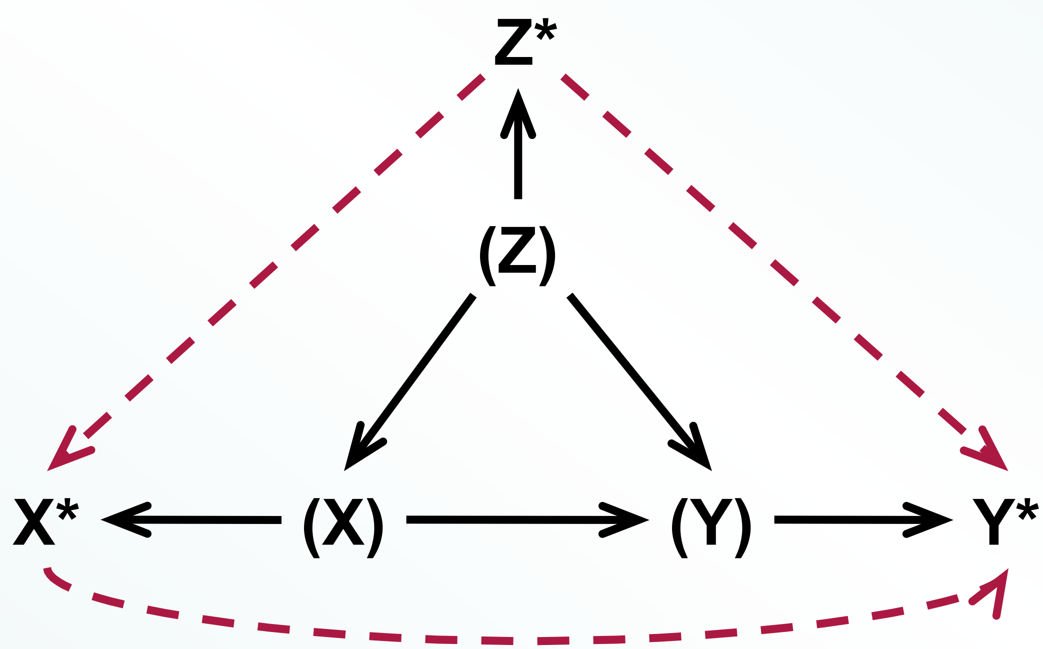




Notice that conditioning on Z* or X* tends to partially block paths containing their true versions provided those are not colliders as seen in the DAGs here.



For example, conditioning on Z* partially blocks the backdoor between (X) and (Y) via Z, thus leaving residual confounding of (X)-(Y).



Independent Differential Measurement Error

First DAG:

Z*-X* association now reflects these biasing paths:

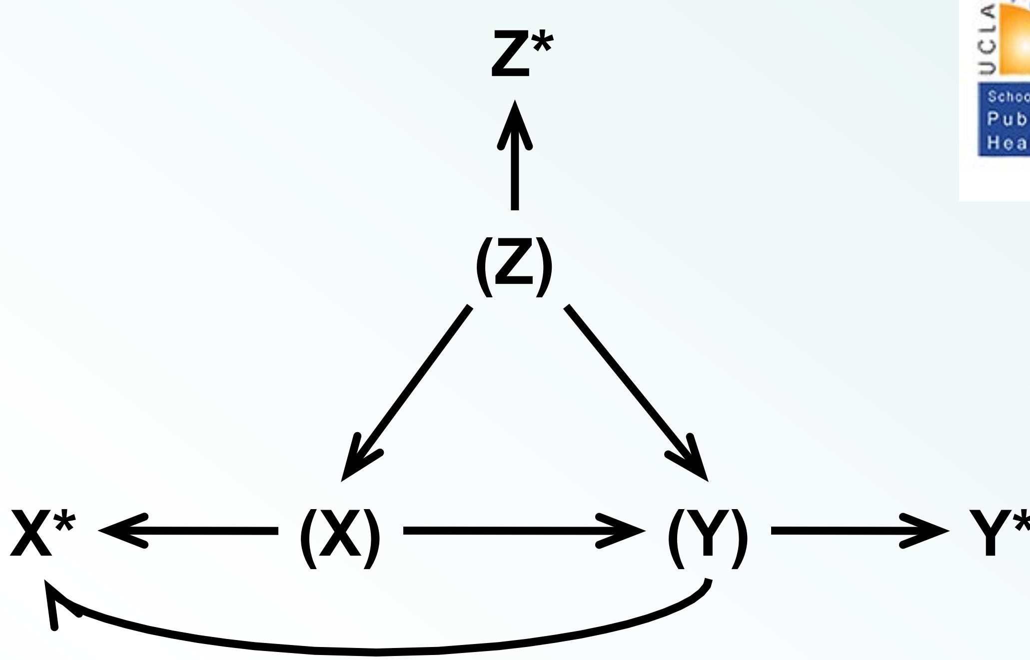
1. Z*-(Z)-(X)-X*
2. Z*-(Z)-(Y)-X*.

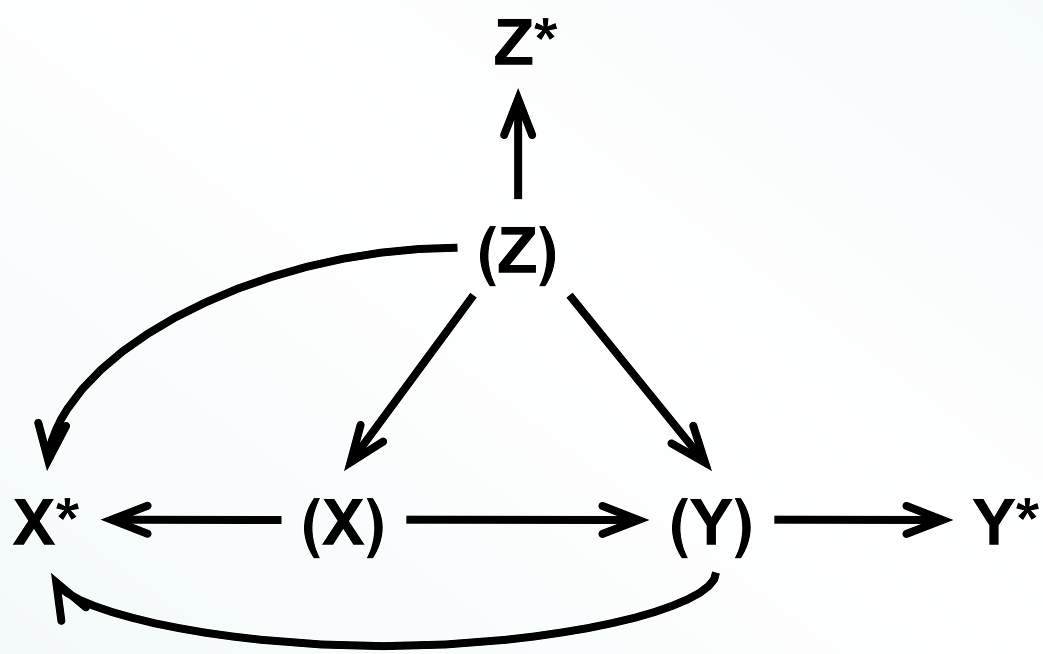
Z*-Y* association:

1. Z*-(Z)-(Y)-Y*
2. Z*-(Z)-(X)-(Y)-Y*
3. Z*-(Z)-(X)-X*-(Y)-Y*.

X*-Y* association:

1. X*-(X)-(Y)-Y*
2. X*-(X)-(Z)-(Y)-Y*
3. X*-(Y)-Y*.



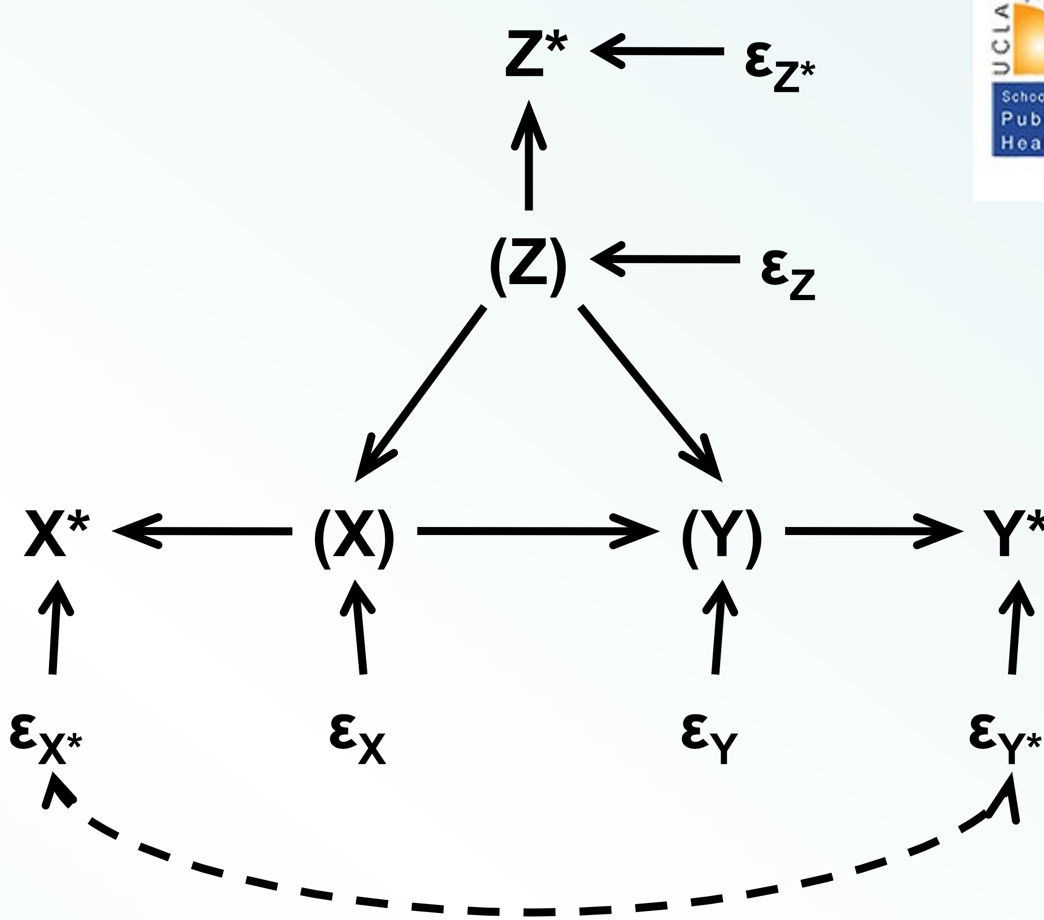


Dependent Non-differential Measurement Error

X^*-Y^* association now reflects these biasing paths:

1. X^*-(X)-(Y)-Y^*
2. X^*-(X)-(Z)-(Y)-Y^*
3. X^*-ε_{X^*}-ε_{Y^*}-Y^*.

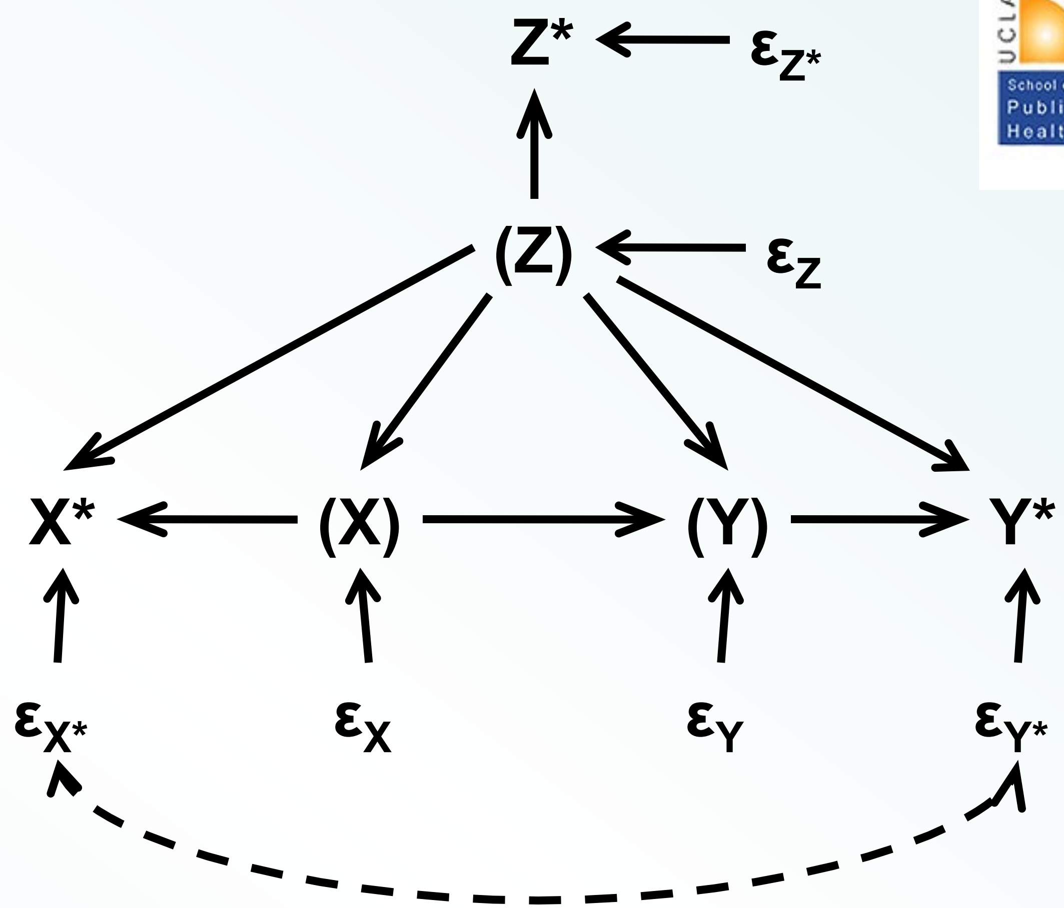
We note that, given this DAG, ε_{X^*} and ε_{Z^*} (or ε_{X^*} and ε_X) are correlated, unlike ε_X and ε_Z, in corresponding multivariable adjustment.



Dependent Differential Measurement Error

X*-Y* association now reflects these biasing paths:

1. X*-(X)-(Y)-Y*
2. X*-(X)-(Z)-(Y)-Y*
3. X*-(Z)-(X)-(Y)-Y*
4. X*-(Z)-(Y)-Y*
5. X*-(Z)-Y*
6. X*-ε_{X^*}-ε_{Y^*}-Y*



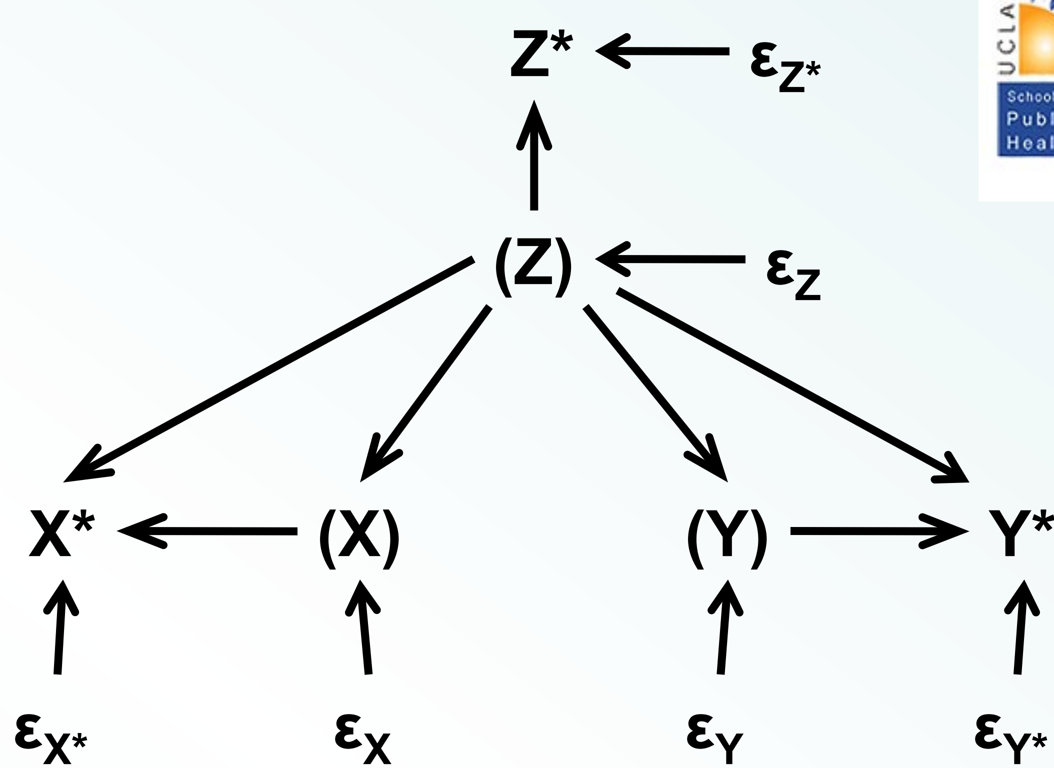
Notable results of ME when true effect is null:

1. ME in confounders
2. Uncontrolled confounding
3. Selection bias
4. Longitudinal (time-varying) data setting

ME in confounder:

If no true X-Y effect, then X*-Y* association will be null if and only if

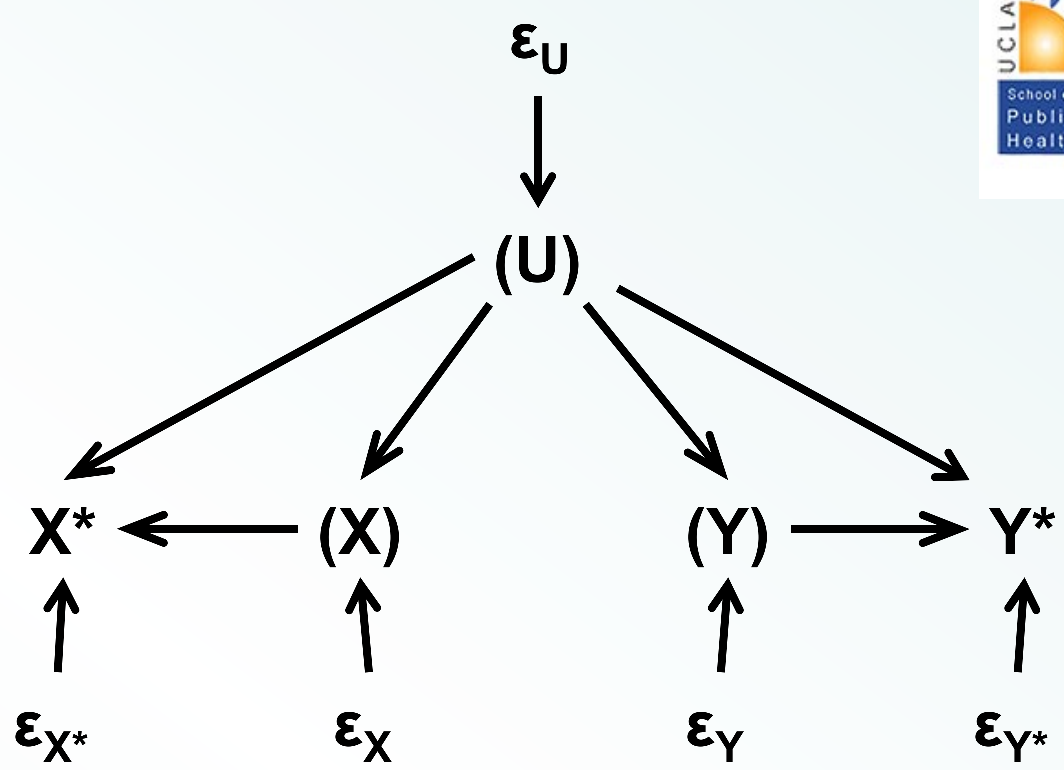
1. No uncontrolled confounding
2. No ME in confounders
3. No dependent ME
4. All causes of differential ME are controlled for, or no differential ME



Uncontrolled confounding:

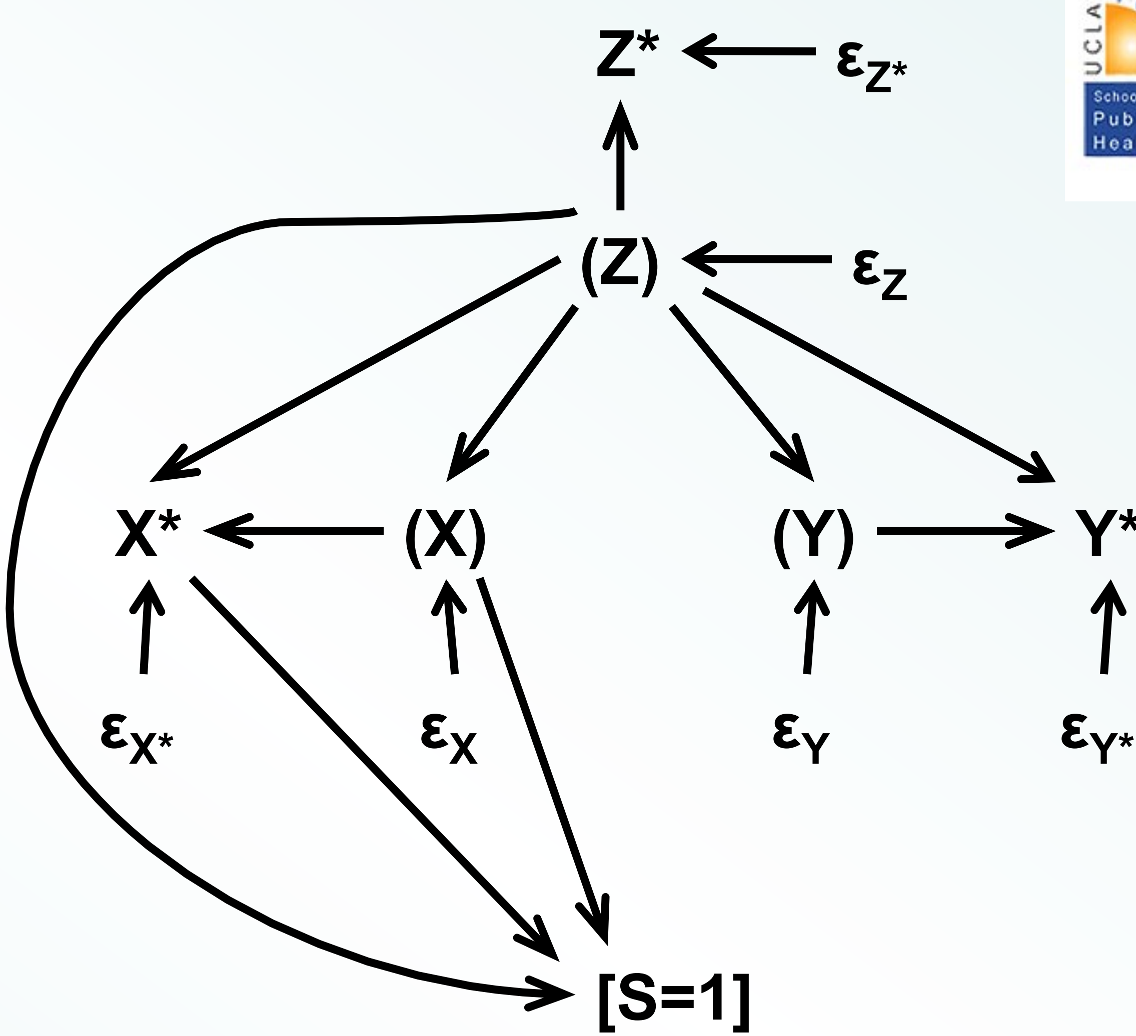
If no true X-Y effect, then X*-Y* association will NOT be null even if

1. **No ME in measured confounders**
2. **No differential ME**
3. **No dependent ME**



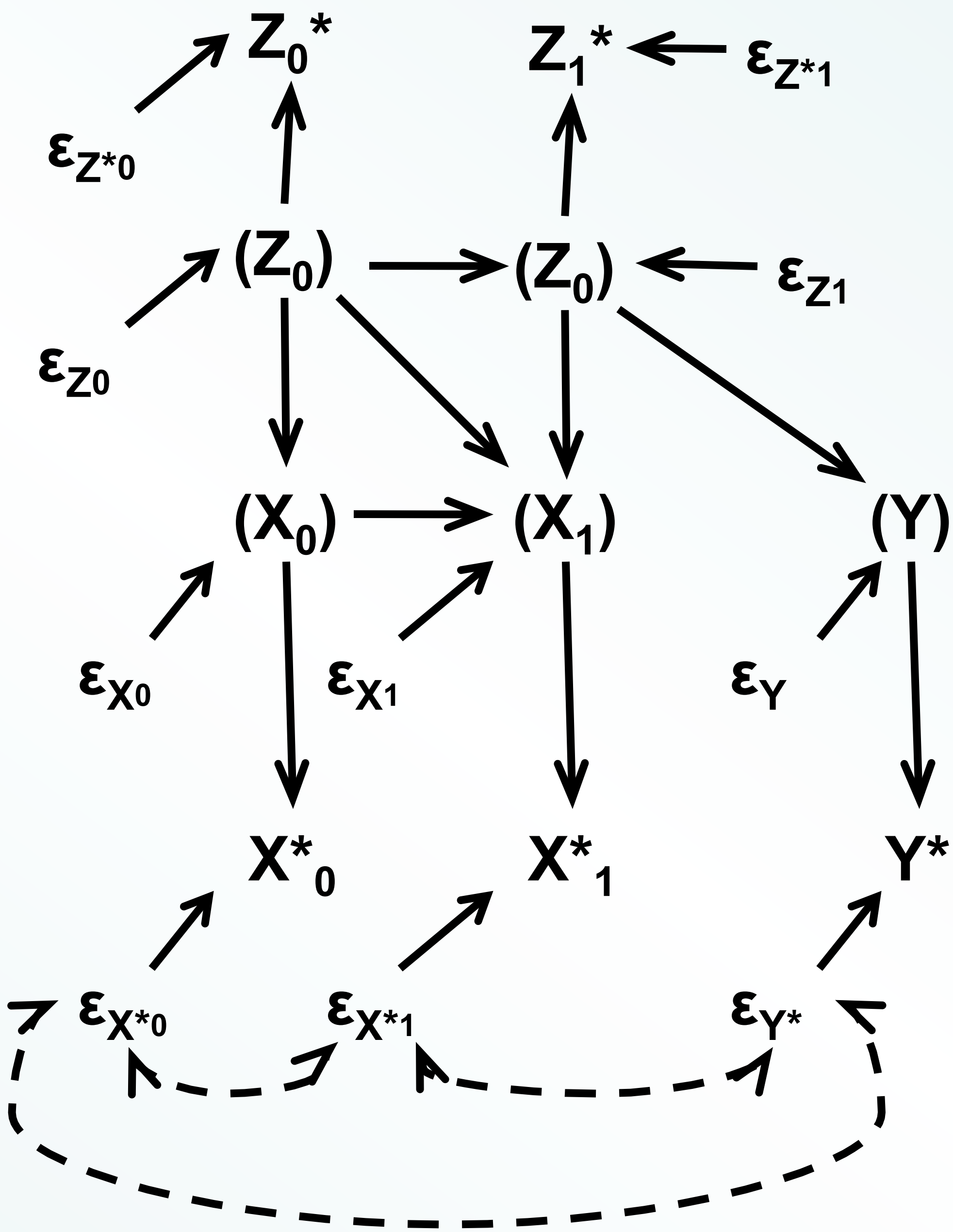
Selection bias (special type):

If no true (X)-(Y) effect, then X*-Y* association will NOT be null given Z* because conditioning on Z* will not completely block the biasing paths allowed by (Z)



Longitudinal (time-varying) data:

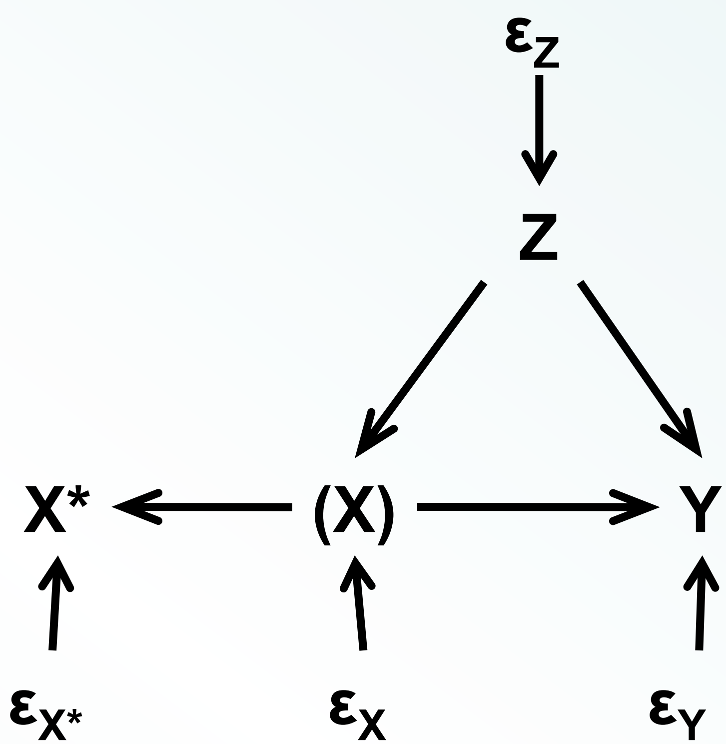
Several biasing paths are present in this simplistic setting even in the absence of effects of true X_0 and X_1 on true Y. Worse when censoring and uncontrolled confounding are built in.



Visualizing Adjustments for Measurement Error

ME as missing data:

Data on true X missing. As can be read from the DAG, to impute X we will need priors for or data on the X*-(X), Z-(X), Y-(X) associations!



$$P(y|x,z)P(x^*|x)P(x|z)P(z) = P(y|x,z,x^*)P(x^*|x,z)P(x|z,)P(z)$$

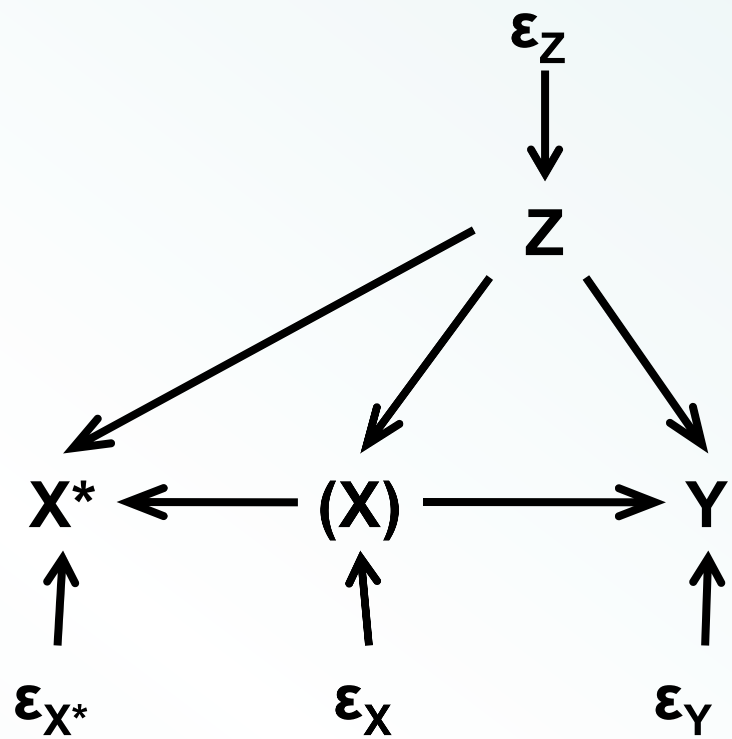
$$\Rightarrow$$

$$P(x|y,z,x^*)P(y,z,x^*)$$

where for a nominal X under a logistic model, $P(x|y,z,x^*)$ can be modeled as the expit of the relevant X*-(X), Z-(X) and Y-(X) coefficients

ME as missing data:

If differential ME as in the DAG here, then in addition to priors for or data on the X*-(X) and Y-(X) associations, we will need the combined Z-X*-(X) and Z-(X) association to impute X because conditioning on X* will open up the extra path Z-X*-(X) between Z and X, beyond the direct Z-(X) relation.



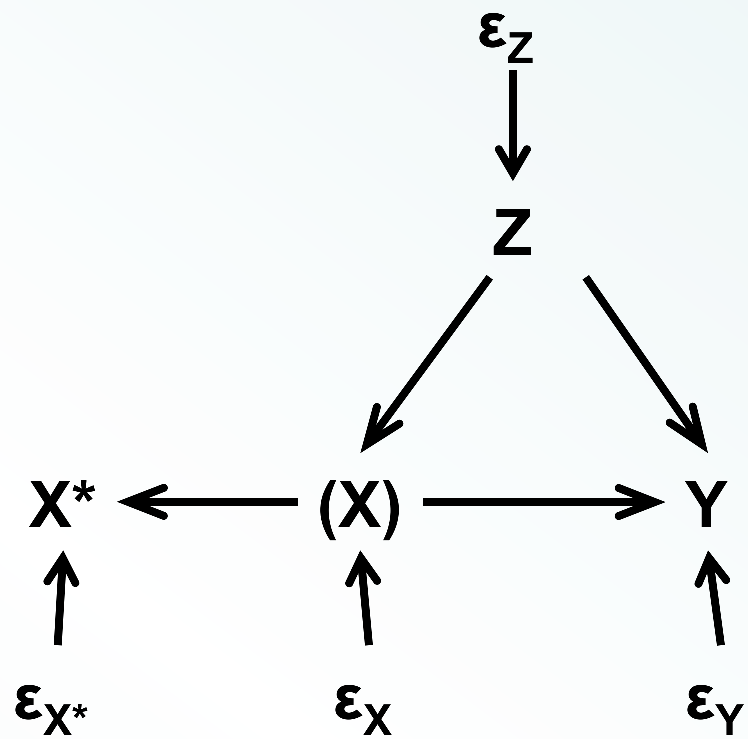
Regression calibration:

X* can be seen as an instrumental variable for X even without satisfying the (causal) instrumentality requirements

$$E(Y|x,z)$$
$$= E(Y|E(X|x^*,z),z)$$
$$= E(Y|E(X^*|x,z),z)$$
$$= E(Y|E(X^*|x),z)$$

but

$$E(Y|x,z)$$
$$\neq E(Y|E(X|x^*),z)$$



Conclusions

Augmented DAGs show that estimated parameters are biased through all paths left partially open by the unobserved true variables.

Augmented DAGs also aid in visualizing the possible methods needed to assess the quantitative impact of ME

Funding Acknowledgement:

Veni Career Grant #916.96.056 from the Netherlands Organization for Scientific Research (NWO)

Special thanks to:

Sander Greenland for helpful comments.