Causal Inference for Health Data

(STATS C160/C260 — Winter 2026)

Lecture 17: Measurement Error

Drago Plečko

UCLA College | Physical Sciences
Statistics & Data Science

What is Measurement Error? When does it occur?

Med Care. 2014 Jun;52(6):e39-43. doi: 10.1097/MLR.0b013e318268ac86.

BMJ Open

Visit this Journal

BMJ

BMJ Open. 2018 Jun 19;8(6):e020857. doi: 10.1136/bmjopen-2017-020857

Circulation

CURRENT ISSUE

REVIEW ARTICLE | Originally Published 30 March 2010 |

Check for updates

Medication Adherence in Cardiovascular Disease

Steven Baroletti, PharmD, MBA, and Heather Dell'Orfano, PharmD | AUTHOR INFO & AFFILIATIONS

Circulation • Volume 121, Number 12 • https://doi.org/10.1161/CIRCULATIONAHA.109.904003

Measurement Error vs. Missing Data

Missing Data

- A value may be unobserved, but we observe R_V, i.e., we know which values are missing

Measurement Error

- A value is observed, but $V^* \neq V$ is possible, and we don't know where the error occurs

True data

Age	Sex	Obesity
16	*F*	*Yes*
15	*M*	*No*

Missing data:

Age	Sex	Obesity
16	***m***	*Yes*
15	*M*	***m***

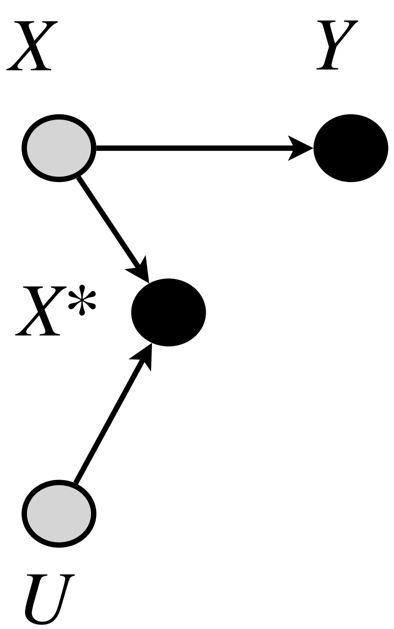
Measurement error:

Age	Sex	Obesity
16	*F*	*Yes*
17	*M*	***Yes***

Types of Measurement Error: Non-differential vs. Differential

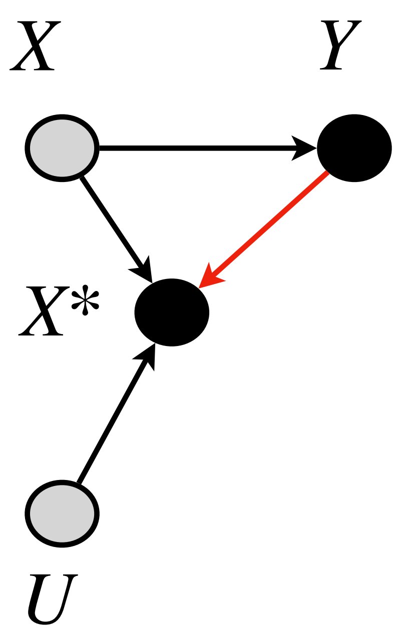
Non-differential

- True Y does not influence how X^* is obtained from X



Differential

- True Y may influence how X^* is obtained from X

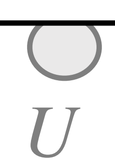


Differential Error may complicate things

Types of Measurement Error: Non-differential vs. Differential

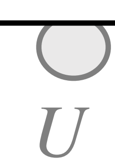
Non-differential

- True Y does not influence how X^* is obtained from X



Differential

- True Y may influence how X^* is obtained from X



Note: Many results in the measurement error literature are for a specific parametric setting.

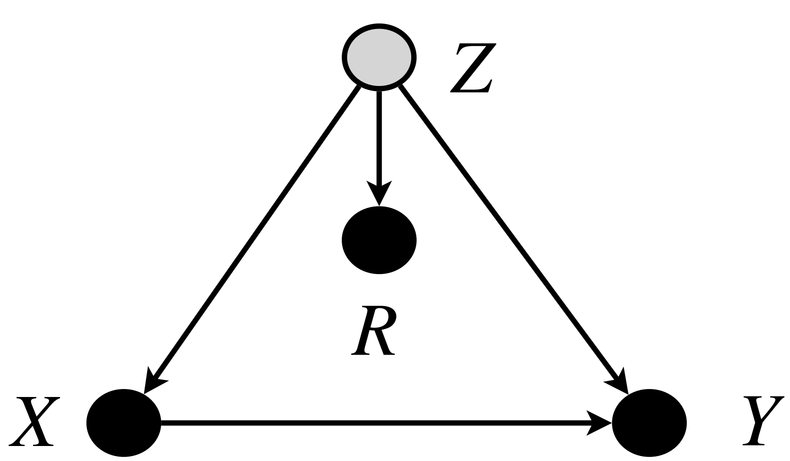
Differential Error may complicate things

Challenges of Measurement Error

- What kind of challenges does measurement error data pose? In the first instance, we can run our analysis (unlike in the missing data setting),
- However, there is a question of whether the obtained results are reliable,
- In this lecture, we will study a specific example of effect identification under measurement error.

Imperfect Confounders

- We consider a setting with treatment X, outcome Y, confounders Z, with the goal of estimating the causal effect of X on Y,
- The confounders, however, are measured imperfectly, and we have access to a proxy R.

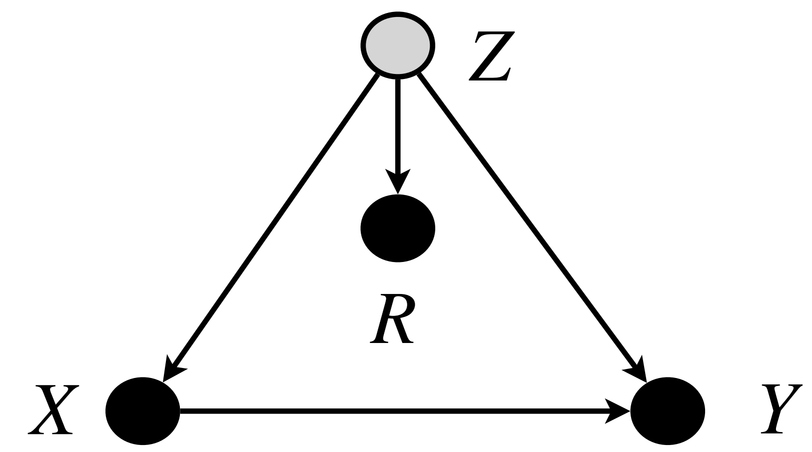


R is not back-door for X, Y — can we somehow identify the effect?

Imperfect Confounders

$$P(y \mid do(x)) = \sum_z P(y \mid z, x) P(z)$$

$$= \sum_z \frac{P(y, z \mid x)}{P(z \mid x)} P(z)$$



$$\Longrightarrow \; R \perp\!\!\!\perp X, Y \mid Z$$

- Note also that we have

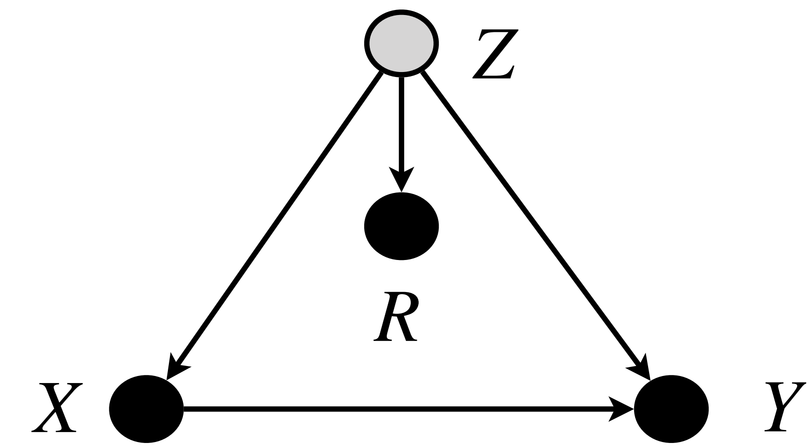
$$P(y, r \mid x) = \sum_z P(y, r, z \mid x)$$

$$= \sum_z P(r \mid x, y, z) P(y, z \mid x)$$

$$= \sum_z P(r \mid z) P(y, z \mid x)$$

Imperfect Confounders

$$P(y, r \mid x) \overset{(*)}{=} \sum_z P(r \mid z) P(y, z \mid x)$$



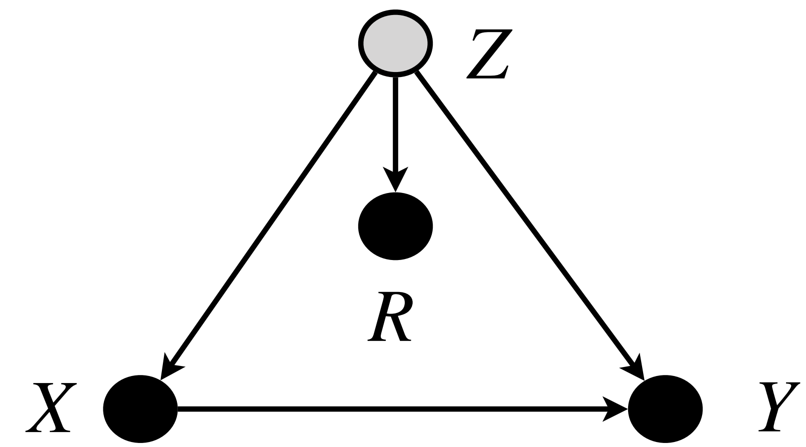
- Suppose that Z, R take a finite number of values,
- We use the following notation:

$$p_{yr|x} = \begin{pmatrix} P(y, r^{(0)} \mid x) \\ \vdots \\ P(y, r^{(|R|-1)} \mid x) \end{pmatrix} \qquad p_{yz|x} = \begin{pmatrix} P(y, z^{(0)} \mid x) \\ \vdots \\ P(y, z^{(|Z|-1)} \mid x) \end{pmatrix}$$

$$A_{r|z} = \begin{pmatrix} P(r^{(0)} \mid z^{(0)}) & \dots & P(r^{(0)} \mid z^{(|Z|-1)}) \\ \vdots & \ddots & \vdots \\ P(r^{(|R|-1)} \mid z^{(0)}) & \dots & P(r^{(|R|-1)} \mid z^{(|Z|-1)}) \end{pmatrix}$$

Imperfect Confounders

$$P(y, r \mid x) \overset{(*)}{=} \sum_z P(r \mid z) P(y, z \mid x)$$



- With this notation, we can re-write $(*)$ as

$$p_{yr|x} = A_{r|z} p_{yz|x}$$

$$\implies \quad p_{yz|x} = A_{r|z}^{-1} p_{yr|x}$$

in words, $P(y, z \mid x)$ values can be obtained if $A_{r|z}$ matrix is known and invertible

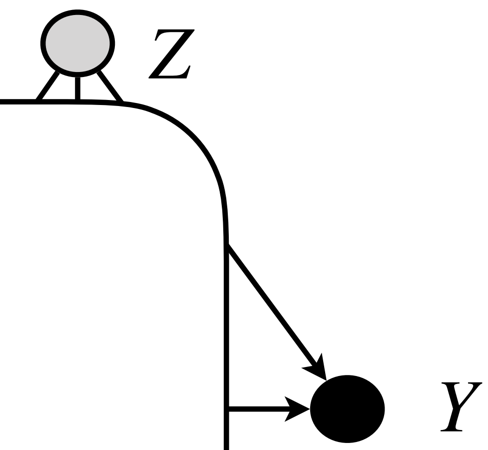
- Similarly, we have that

$$p_{z|x} = A_{r|z}^{-1} p_{r|x}, \; p_z = A_{r|z}^{-1} p_r$$

so $P(z \mid x), P(z)$ recoverable similarly

Imperfect Confounders

$P(y, r$ $(*)$



Since we have

$$P(y \mid do(x)) = \sum_z \frac{P(y, z \mid x)}{P(z \mid x)} P(z)$$

matrix $A_{r|z}$ determines the causal effect!

- With t

values

matrix

ertible

How can one get $A_{r|z}$?

- Simila

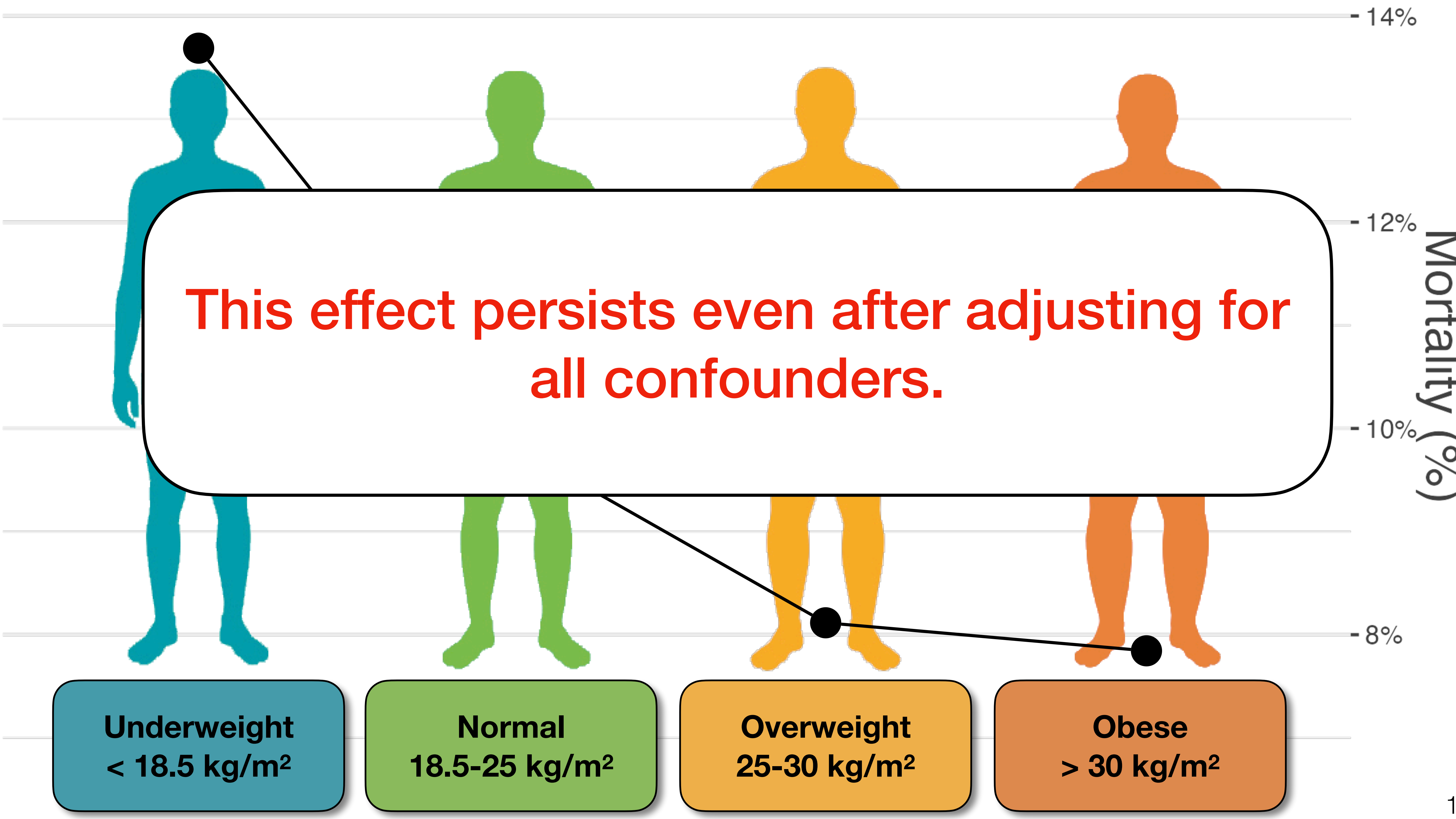
overable

Imperfect Confounders: Help with external data

- Suppose we have access to another dataset, in which the confounders Z and the proxy R are recorded simultaneously,
- This means we have access to samples from the joint distribution $P(z, r)$, meaning that the matrix $A_{r|z}$ can be determined,
- This is known as recovery by external data, which is commonly needed as measurement error can be difficult to solve.

Application: Obesity Paradox

- For patients in intensive care units, we observe:



Obesity Paradox: Explanations

- There are two types of explanations of the OP

Causal

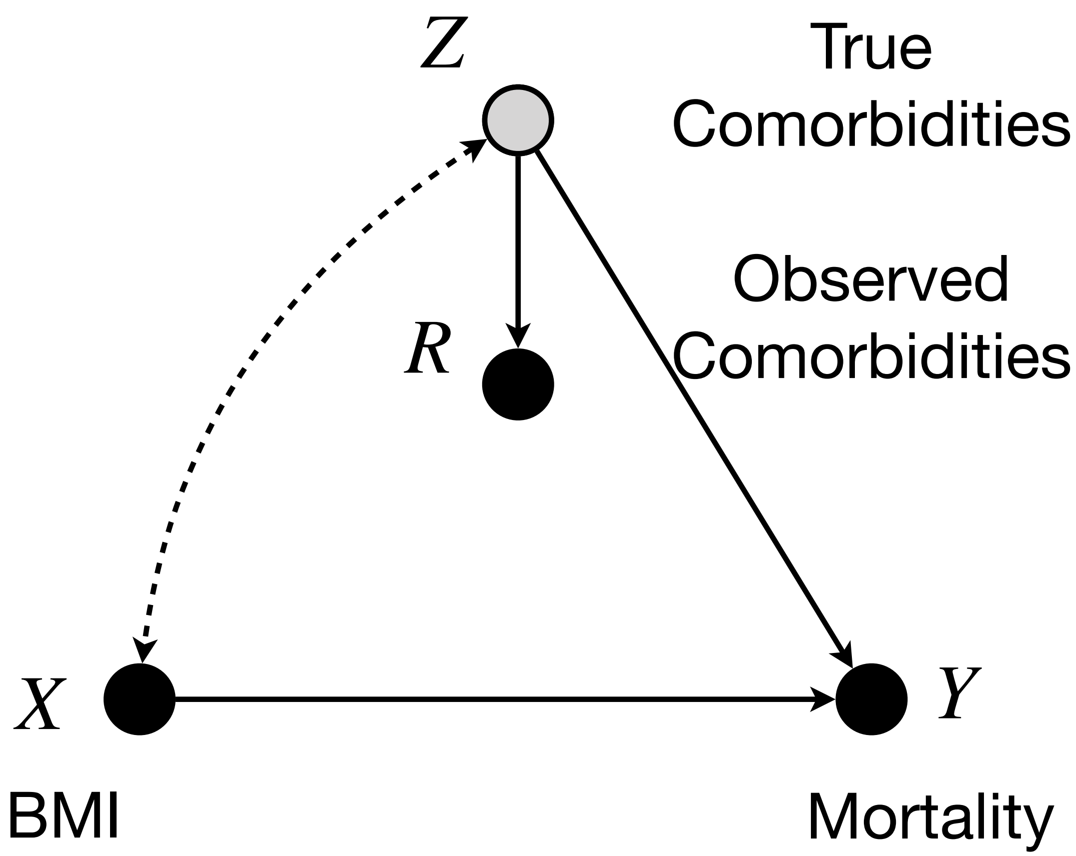
- increased body mass index —> increased fat issue, and adipose tissue may absorb inflammation (16998510)
- increased body mass index —> increased muscle mass, and muscle may serve as physiological reserve of protein/energy

muscle loss 2%/day in ICU (36597123)

Confounded

- decreased body mass index may be associated with worse chronic health, since comorbidities such as cancer/smoking cause weight loss
- if chronic health captured imperfectly, we may have residual confounding

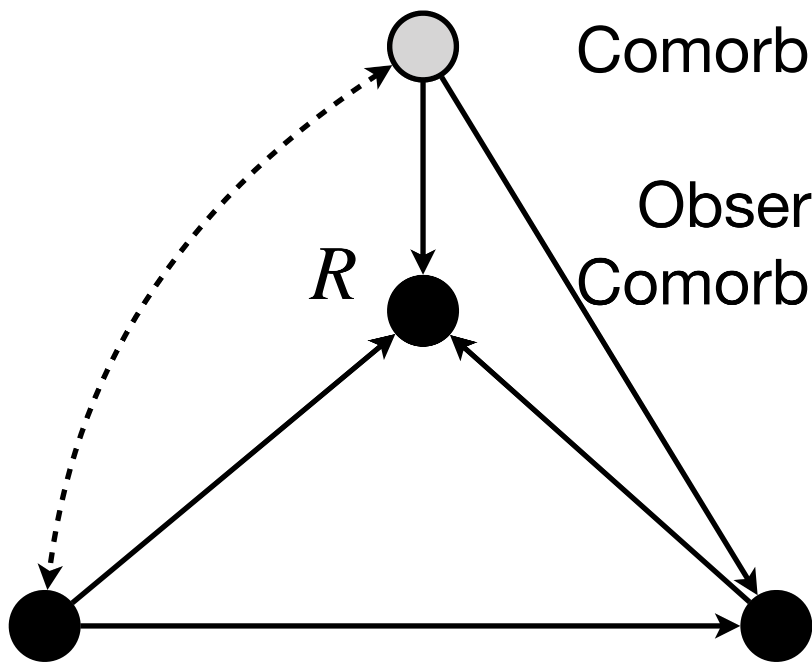
Obesity Paradox: A Sensitivity Approach



Under-reporting of comorbidities is a common concern — and may be a source of measurement error.

Question: how much measurement error is needed to flip the causal effect estimate?

Obesity Paradox: Model of Measurement Error (Univariate)



- Suppose that we have

$$R \leftarrow \begin{cases} \textbf{Bernoulli}(\phi') & \textbf{if } Z = 0 \\ \textbf{Bernoulli}(1 - \phi) & \textbf{if } Z = 1, \end{cases}$$

- Provided the values ϕ, ϕ', the conditional $P(R \mid Z)$ is determined.

Given values $\phi, \phi', P(y \mid do(x))$ can be obtained using the Kuroki-Pearl result

But what if $X \rightarrow R, Y \rightarrow R$? Differential errors are hard 😞

Obesity Paradox: Effect Recovery

Theorem. The average treatment effect of X on Y can be uniquely computed from the joint distribution $P(r, x, y)$ provided matrices $A_{r|xyz}$, if one of the following two conditions holds:

- The no false positive assumption holds, $R_m \leq Z_m$ for every component $m \in \{1, \ldots, k\}$, and $P(R = z^{(i)} \mid Z = z^{(i)}, x, y) > 0 \; \forall \; z^{(i)}, x, y$
- The stability assumption holds, $P(R = z^{(i)} \mid Z = z^{(i)}, x, y) > \frac{1}{2} \; \forall \; z^{(i)}, x, y \, .$

Effect Recovery: Proof

- Note that the following identify holds:

$$\underbrace{\begin{pmatrix} \vdots \\ P(r^{(i)} \mid x, y) \\ \vdots \end{pmatrix}}_{p_{r|xy}} = \underbrace{\begin{pmatrix} \ddots & \vdots & \vdots \\ \vdots & P(r^{(i)} \mid z^{(j)}, x, y) & \vdots \\ \vdots & \vdots & \ddots \end{pmatrix}}_{A_{r|xyz} \text{ (size } 2^k \times 2^k)} \underbrace{\begin{pmatrix} \vdots \\ P(z^{(j)} \mid x, y) \\ \vdots \end{pmatrix}}_{p_{z|xy}}$$

- Therefore, if $A_{r|xyz}$ is invertible, we can recover $P(z \mid x, y)$ from $P(r \mid x, y)$.

Effect Recovery: Proof

- Suppose that $P(z \mid x, y)$ is available for each x, y.

$$P(y \mid do(x)) = \sum_z P(y \mid z, x) P(z)$$

weights $P(x, y)$ are observed!

$$= \sum_z \frac{P(y, x, z)}{P(x, z)} P(z)$$

we only need to show $A_{r|xyz}$ invertibility

$$= \sum_z \frac{P(y, x, z)}{\sum_{y'} P(y', x, z)} \sum_{x', y'} P(y', x', z)$$

$$= \sum_z \frac{P(z \mid x, y)}{\sum_{y'} P(z \mid y', x) P(y', x)} \sum_{x', y'} P(z \mid y', x') P(y', x')$$

Effect Recovery: Proof

- Suppose that $P(z \mid x, y)$ is available for each x, y.

$$P(y \mid do(x)) = \sum_z P(y \mid z, x) P(z)$$

weights $P(x, y)$ are observed!

We prove invertibility under (i) no false positives; (ii) stability on the white board!

$$= \sum_z \frac{P(z \mid x, y)}{\sum_{y'} P(z \mid y', x) P(y', x)} \sum_{x', y'} P(z \mid y', x') P(y', x')$$

Simplifying Our Sensitivity Analysis: Assumptions

- Specifying the matrix $A_{r|xyz}$ requires 2^{2k} parameters, which is huge in practice.
- We make some simplifying assumptions:

No False Positives (NFP)

$$R_m \leq Z_m,$$

any comorbidity not present ($Z_m = 0$) will be recorded as not present

Independent Fidelity (IF)

$$P(r \mid x, y, z) = \prod_i P(r_i \mid x, y, z_i)$$

Measurement Error indep. across comorbidities

Parameter Sharing (PS)

$$P(r_m \mid z_m, x, y) = P(r_\ell \mid z_\ell, x, y) = \phi_{xy}$$

equal error rates for each variable

ϕ-value Definition

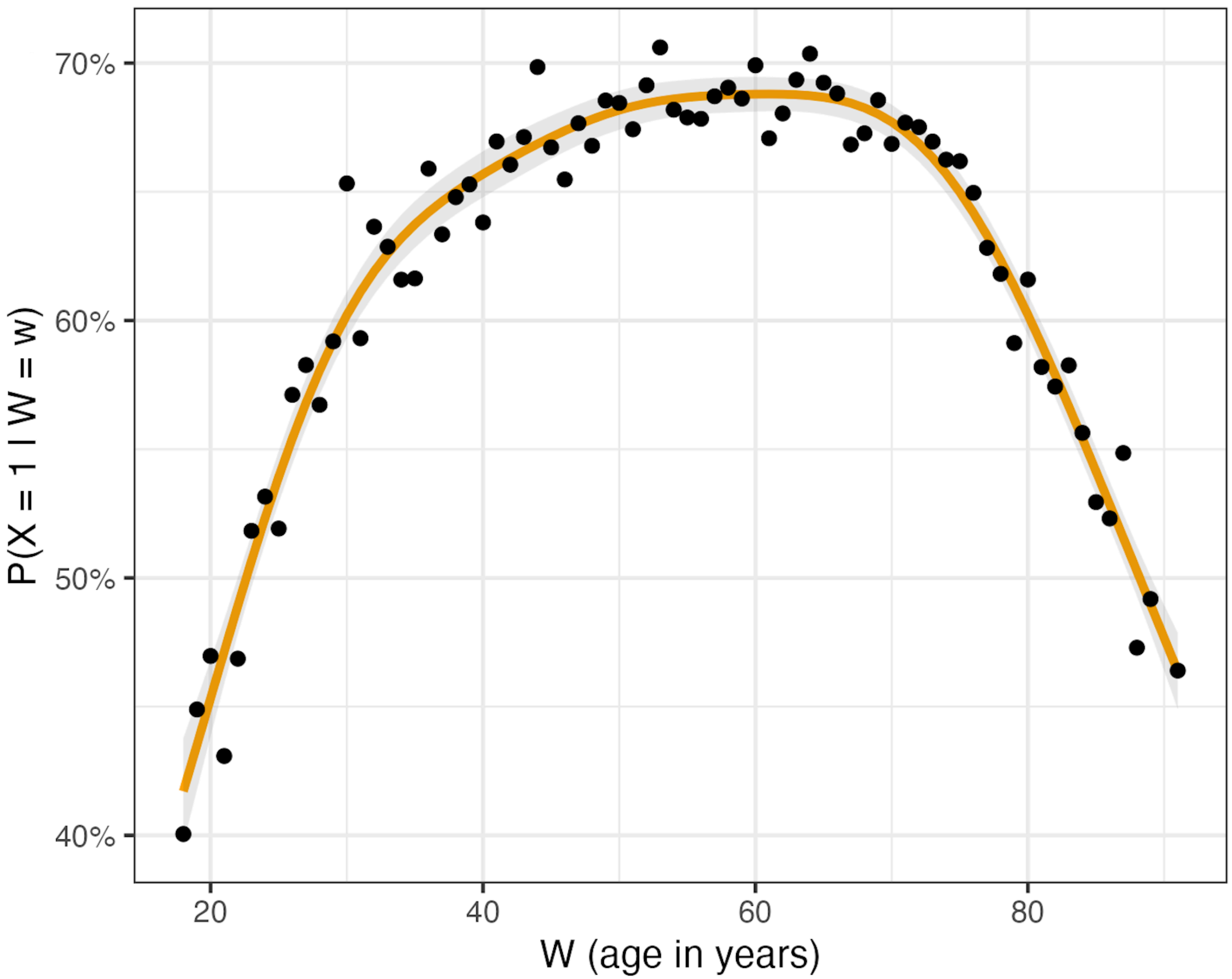
Definition. A ϕ-value is any fidelity pattern $\Phi_{crit} = (\phi_{x_0y_0}, \phi_{x_0y_1}, \phi_{x_1y_0}, \phi_{x_1y_1})$ such that the under Φ the causal effect estimate changes sign compared to the no missingness setting $\Phi = (0,0,0,0)$.

$$\text{TE}_{x_0,x_1}(y \mid \Phi = 0) * \text{TE}_{x_0,x_1}(y \mid \Phi_{crit}) < 0.$$

Intuition: We are searching for a minimal amount of measurement error that flips the sign!

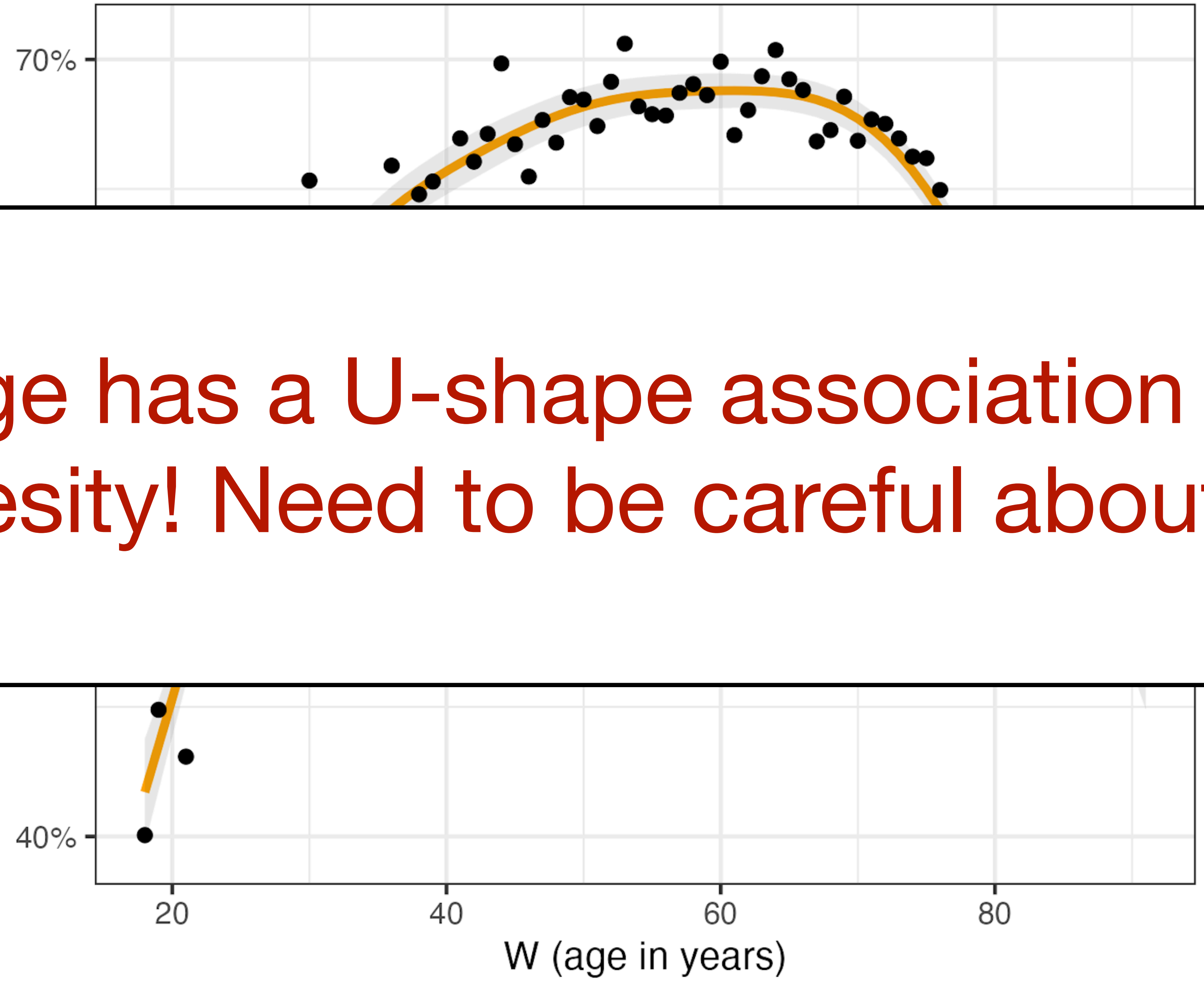
ϕ-value: Inference

- In our setting, we have an important continuous confounder W which is measured without error, age



ϕ-value: Inference

- In our setting, we have an important continuous confounder W which is measured without error, age



Age has a U-shape association with obesity! Need to be careful about this.

ϕ-value: Inference

- We therefore use a parametric model as follows:

$$P(z, w) = \frac{1}{A(\Sigma, \Lambda, \Omega, b)} \exp(z^T \Sigma z + w^T \Lambda z - \frac{1}{2} w^T \Omega w + b^T w)$$

quadratic term in W!

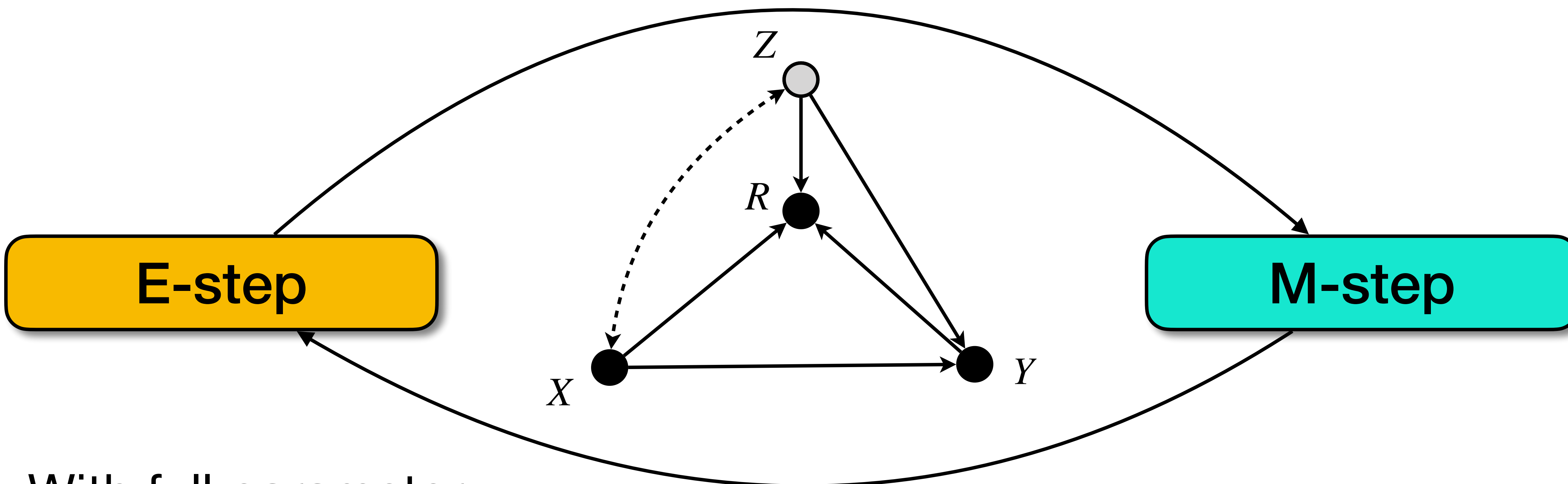
$$P(X = 1 \mid Z = z, W = w) = \text{expit}(\lambda_z^T (1, z) + \lambda_w^T (w, w^2))$$

$$P(Y = 1 \mid Z = z, X = x) = \text{expit}(\mu_z^T (1, z) + \mu_w w + \beta x)$$

parameter of interest

Under Assumptions NFP+IF+PS we are searching for a minimal ϕ value

ϕ-value: Inference via Expectation-Maximization Intuition



- With full parameter specification $\theta^{(t)}$, we know how to sample $P_{\theta^{(t)},\Phi}(z \mid x, y, r)$,
- Allows us to create a "full-sample" as if Z was observed.

Alternate steps until convergence

- With full data (including Z) we know how to estimate $\theta^{(t)}$,
- Allows us to update our parameters to $\theta^{(t+1)}$.

Obesity Paradox Measurement Error: Discussion of Assumptions

- Recall some of the assumptions we used:

No False Positives (NFP)

$$R_m \leq Z_m$$

It is rather unlikely that a comorbidity that does not exist will be recorded by mistake — this assumption is strongly supported.

Independent Fidelity (IF)

$$P(r \mid x, y, z) = \prod_i P(r_i \mid x, y, z_i)$$

Depends on data collection.
If comorbidities are recorded by different departments, strong justification.
If a single ICD-coder, less likely.

Parameter Sharing (PS)

$$P(r_m \mid z_m, x, y) = P(r_\ell \mid z_\ell, x, y) = \phi_{xy}$$

Depends on data collection.
If comorbidities are recorded by different departments, may be less likely.
If a single ICD-coder, more likely.

Obesity Paradox Measurement Error: Discussion of Assumptions

- Recall some of the assumptions we used:

No False Positives (NFP)

$$R_m \leq Z_m$$

It is rather unlikely that a comorbidity that does not exist will be recorded by mistake — this assumption is strongly supported.

Independent Fidelity (IF)

Depends on data collection.

Note: IF and PS assumptions are only really needed to *simplify the specification of the ME pattern* — they are not needed for inference (i.e., we can remove them)

If a single ICD-coder, more likely.