Causal Inference for Health Data

(STATS C160/C260 — Winter 2026)

Lecture 5: The Problem of Confounding and the Back-door Criterion

Drago Plečko

UCLA College | Physical Sciences
Statistics & Data Science

Recap: do(x) World

$M_{\text{real}} \quad \Longrightarrow \quad do(Smoke = t) \quad M_{\text{Smoke=t}}$

$Smoke \leftarrow \mathrm{f}_{\text{Smoke}}(U_s,\ U_g)$

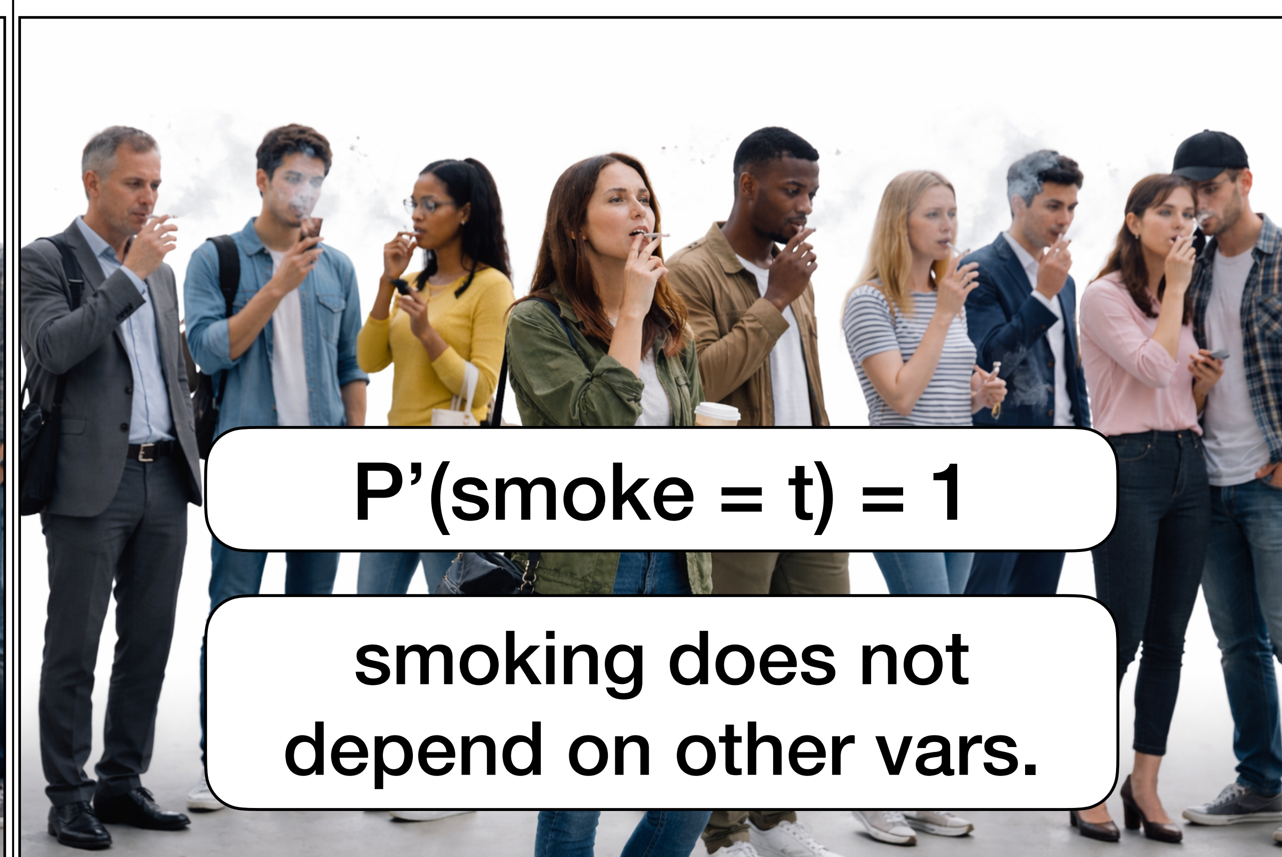
$Cancer \leftarrow \mathrm{f}_{\text{Cancer}}(Smoke,\ U_c,\ U_g)$

$\Downarrow \quad P$

$Smoke \leftarrow \text{true}$

$Cancer \leftarrow \mathrm{f}_{\text{Cancer}}(Smoke,\ U_c,\ U_g)$

$\Downarrow \quad P'$ (hypothetical)



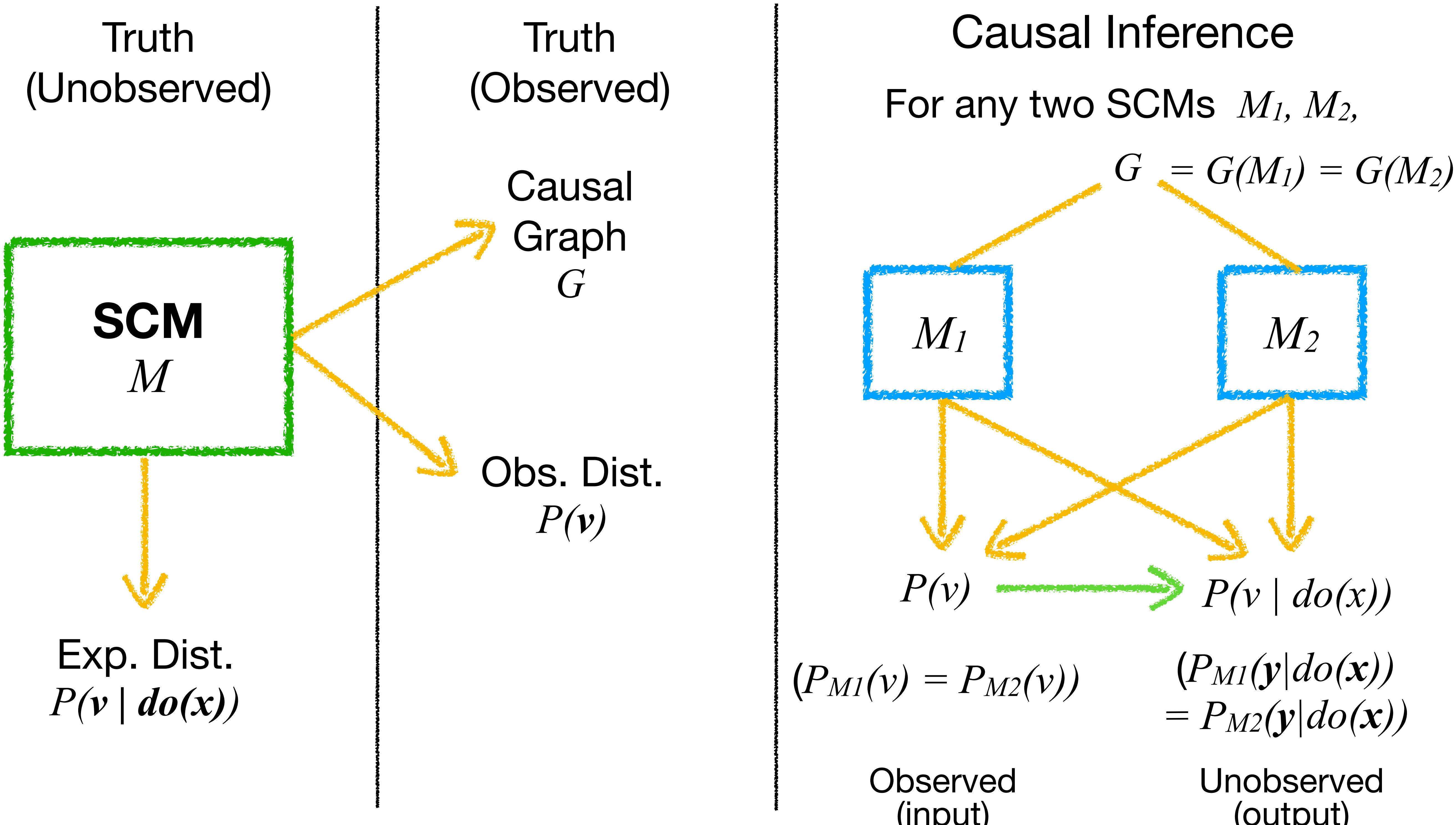
Recap:
The Identification Problem

Causal Effect Identifiability (Def. 4.2.1)

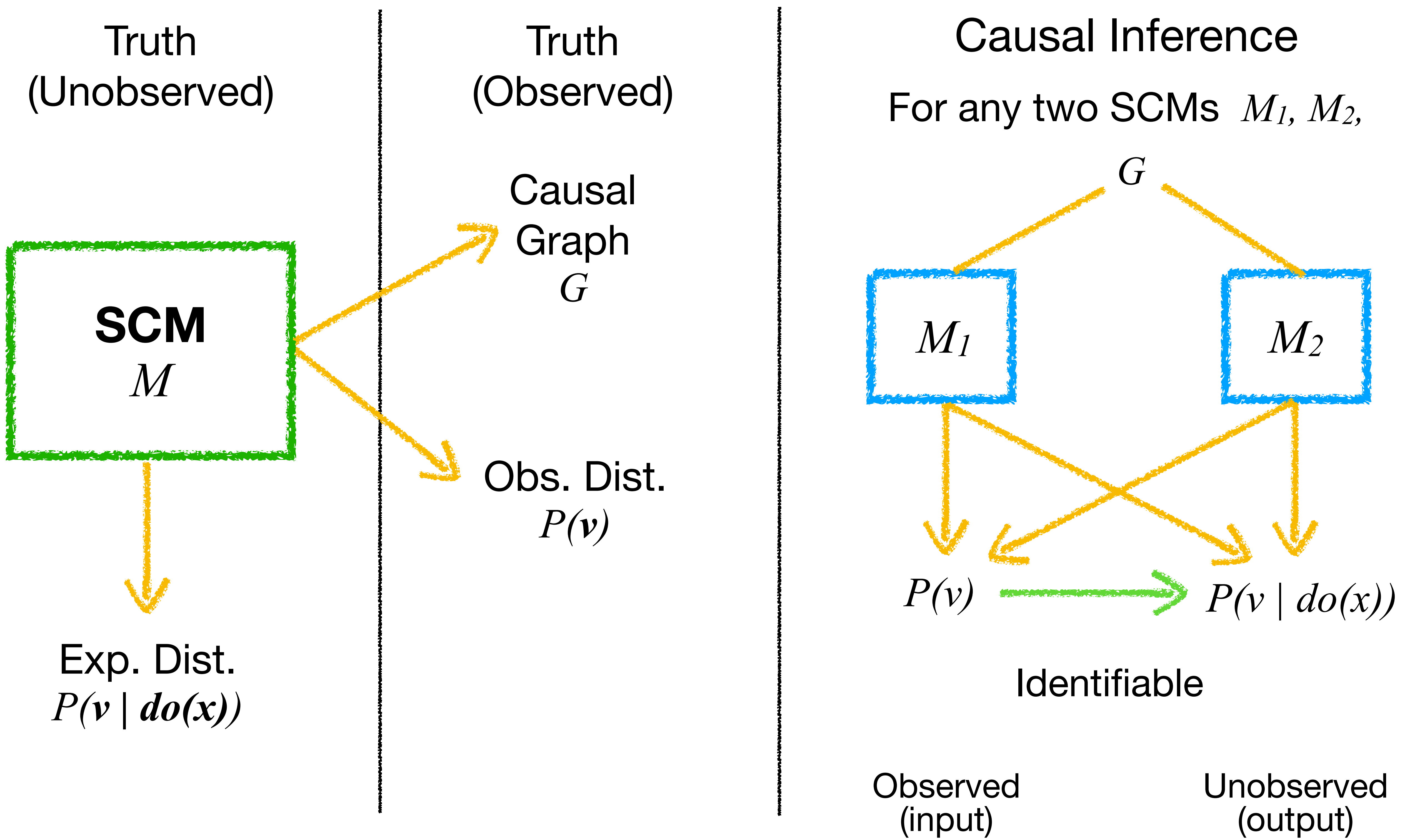
The causal effect of $\boldsymbol{X}$ on $\boldsymbol{Y}$ is said to be identifiable from a causal diagram G if the quantity $P(\mathbf{y} \mid do(\mathbf{x}))$ can be computed uniquely from a positive probability of the observed variables.

That is, if for every pair of models M_1 and M_2 inducing G, $P_{M_1}(\mathbf{y} \mid do(\mathbf{x})) = P_{M_2}(\mathbf{y} \mid do(\mathbf{x}))$, whenever $P_{M_1}(\mathbf{v}) = P_{M_2}(\mathbf{v}) > 0$.

The Identification Problem (II)



The Identification Problem (II)



The Identification Problem (II)

Truth (Unobserved)

Truth (Observed)

Causal Inference

For any two SCMs M_1, M_2,

Identifiability really means that, no matter the shape of M_1, M_2, for all models agreeing in terms of $\langle G, P(\mathbf{v})\rangle$, they will also agree in $P(\mathbf{v} \mid do(\mathbf{x}))$!

SCM
M

M_2

Exp. Dist.
$P(\mathbf{v} \mid \mathbf{do(x)})$

$P(v)$ → $P(v \mid do(x))$

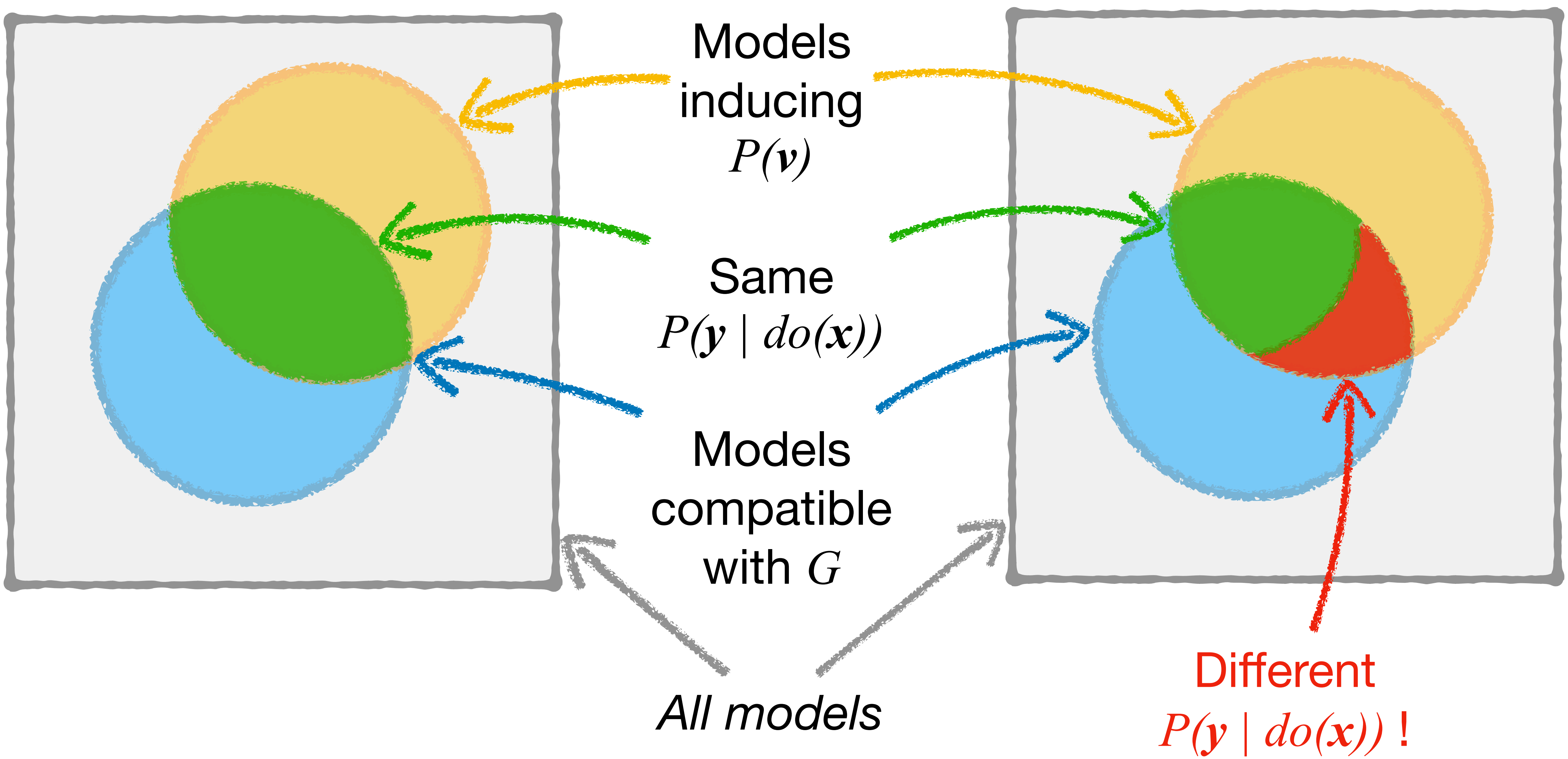
Identifiable

Observed (input)

Unobserved (output)

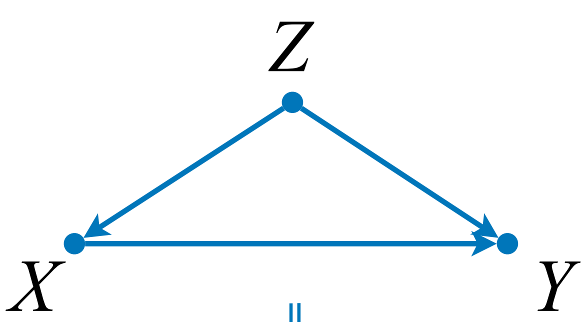
The Identification Problem (III)

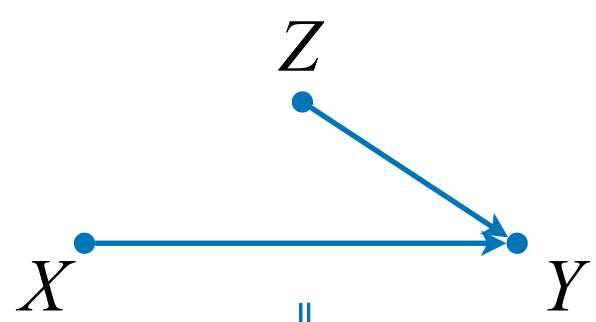
identifiable vs not identifiable



Example. Identifiable Effect

- Consider any two pair of models compatible with the following graph and the same observational distribution $P(\mathbf{v})$:





$$\Downarrow \qquad\qquad \Downarrow$$

$$P(\mathbf{v}) = P(z)P(x|z)P(y|x,z) \qquad P(\mathbf{v}|do(x)) = P(z)P(y|x,z) \implies P(y \mid do(x)) = \sum P(z)P(y \mid x, z)$$

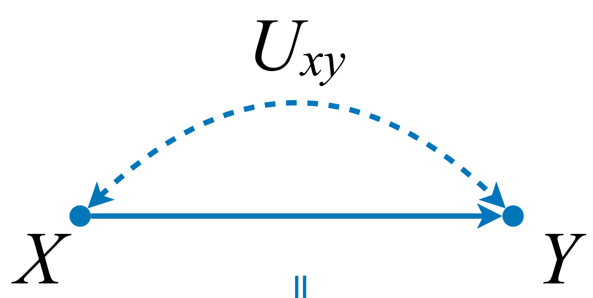
$$M^{(1)} = \begin{cases} Z \leftarrow f_z^{(1)}(u_z) \\ X \leftarrow f_x^{(1)}(z, u_x) \\ Y \leftarrow f_y^{(1)}(x, z, u_y) \end{cases} \quad \overset{do(x)}{\longmapsto} \quad M^{(1)} = \begin{cases} Z \leftarrow f_z^{(1)}(u_z) \\ X \leftarrow x \\ Y \leftarrow f_y^{(1)}(x, z, u_y) \end{cases}$$

$$M^{(2)} = \begin{cases} Z \leftarrow f_z^{(2)}(u_z) \\ X \leftarrow f_x^{(2)}(z, u_x) \\ Y \leftarrow f_y^{(2)}(x, z, u_y) \end{cases} \qquad M^{(2)} = \begin{cases} Z \leftarrow f_z^{(2)}(u_z) \\ X \leftarrow x \\ Y \leftarrow f_y^{(2)}(x, z, u_y) \end{cases}$$

No matter what the specific functions or P(u) are, as long as M_1, M_2 agree in $\langle G, P(\mathbf{v})\rangle$, they will also agree in $P(z)$ and $P(y|x,z)$, hence in $P(\mathbf{v} \mid do(\mathbf{x}))$!

Example. Non-identifiable Effect

- Consider the pair of models compatible with the following graph G and observational distribution $P(\mathbf{v})$:



$\Downarrow$

$$P(\mathbf{v}) = \sum_{u_{xy}} P(y \mid x, u_{xy}) P(x \mid u_{xy}) P(u_{xy})$$

$$M^{(1)} = \begin{cases} X \leftarrow U_{xy} \\ Y \leftarrow (X \oplus U_{xy}) \vee U_y \end{cases}$$

$$M^{(2)} = \begin{cases} X \leftarrow U_{xy} \\ Y \leftarrow U_y \end{cases}$$

$$P^{(j)}(U_i = 1) = 1/2, \quad i = \{y, xy\}, j = \{1,2\}$$

U_y	U_{xy}	X	$Y^{(1)}$	$Y^{(2)}$	$P(\mathbf{v},\mathbf{u})$
0	0	0	0	0	1/4
0	1	1	0	0	1/4
1	0	0	1	1	1/4
1	1	1	1	1	1/4

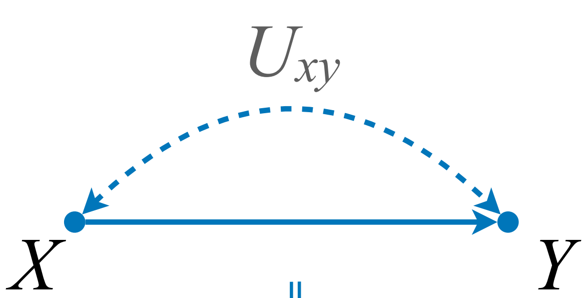
$\Downarrow$

X	$Y^{(1)}$	$Y^{(2)}$	$P^{(1)}(\mathbf{v})$	$P^{(2)}(\mathbf{v})$
0	0	0	1/4	1/4
0	1	1	1/4	1/4
1	0	0	1/4	1/4
1	1	1	1/4	1/4

They match in $P(\mathbf{v})$, that is, $P^{(1)}(\mathbf{v})=P^{(2)}(\mathbf{v})$!

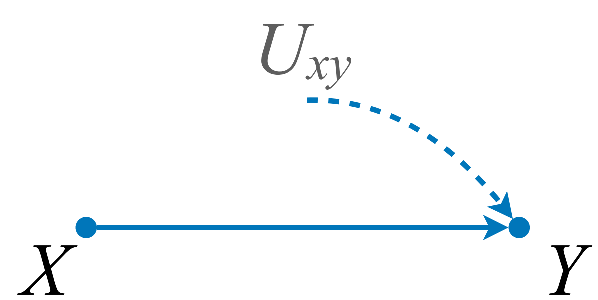
Example. Non-identifiable Effect

- Consider the pair of models compatible with the following graph G and observational distribution $P(\mathbf{v})$:



$$\Downarrow$$

$$P(\mathbf{v}) = \sum_{u_{xy}} P(y \mid x, u_{xy}) P(x \mid u_{xy}) P(u_{xy})$$



$$P(\mathbf{v} \mid do(x)) = \sum_{u_{xy}} P(y \mid x, u_{xy}) P(u_{xy})$$

U_y	U_{xy}	X	$Y^{(1)}$	$Y^{(2)}$	$P(\mathbf{v},\mathbf{u} \mid do(x))$
0	0	x	$0 \oplus x$	0	1/4
0	1	x	$1 \oplus x$	0	1/4
1	0	x	1	1	1/4
1	1	x	1	1	1/4

$$\Downarrow$$

$Y^{(1)}$	$Y^{(2)}$	$P^{(1)}(\mathbf{y} \mid do(x))$	$P^{(2)}(\mathbf{y} \mid do(x))$
0		1/4	1/2
1		3/4	1/2

$$M^{(1)} = \begin{cases} X \leftarrow U_{xy} \\ Y \leftarrow (X \oplus U_{xy}) \vee U_y \end{cases} \qquad M^{(1)} = \begin{cases} X \leftarrow x \\ Y \leftarrow (X \oplus U_{xy}) \vee U_y \end{cases}$$

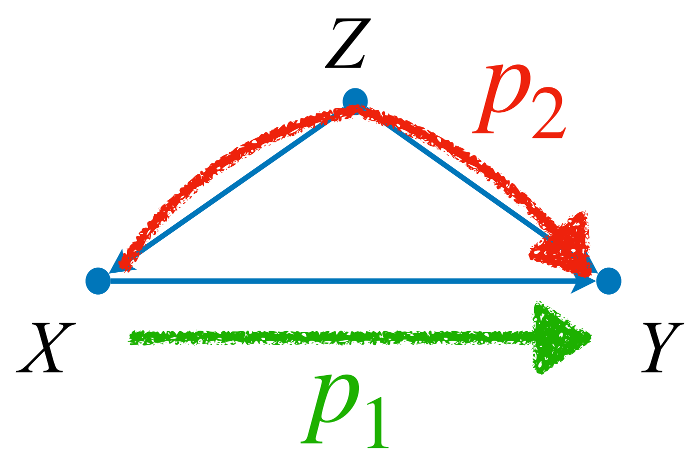
$$\xrightarrow{do(x)}$$

$$M^{(2)} = \begin{cases} X \leftarrow U_{xy} \\ Y \leftarrow U_y \end{cases} \qquad M^{(2)} = \begin{cases} X \leftarrow x \\ Y \leftarrow U_y \end{cases}$$

$$P^{(j)}(U_i = 1) = 1/2, \quad i = \{y, xy\}, j = \{1,2\}$$

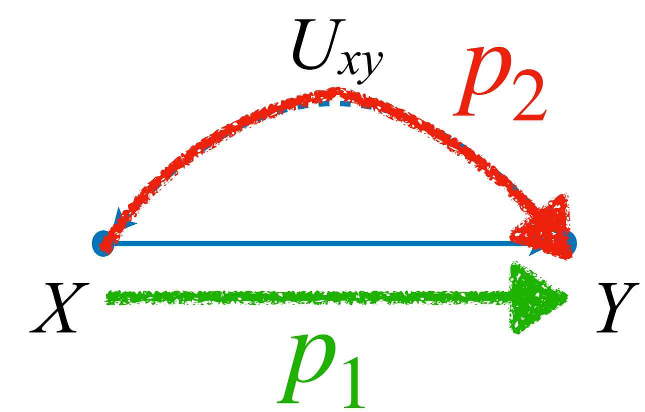
Even though both models induce G and have the same $P(\mathbf{v})$, the effect $P^{(1)}(y \mid do(x)) \neq P^{(2)}(y \mid do(x))$!

Intuition: Non-ID vs. ID



- there is a flow of information from X to Y along two different pathways,
- causal pathway p_1,
- spurious (or confounded) pathway p_2,
- to block p_2, we condition on $Z = z$ and average over it:

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



- same flow of information as in the RHS,
- causal pathway p_1,
- spurious (or confounded) pathway p_2,
- however, we do not see the latent U_{xy}, and therefore cannot block p_2,
- this makes the effect non-identifiable, since we cannot separate p_1 from p_2

Let’s study how to *decide* whether a causal effect is *identifiable*…

Identification in Markovian Models

Theorem 4.2.1. Given the causal diagram G of any Markovian model that all variables are measured, the causal effect $Q = P(\mathbf{y} \mid do(\mathbf{x}))$ is identifiable for every subsets of variables $\boldsymbol{X}$ and $\boldsymbol{Y}$ and is obtained from the truncated factorization,

$$P(\mathbf{v} \mid do(\mathbf{x})) = \prod_{\{V_i \in \mathbf{V} \setminus \mathbf{X}\}} P\left(v_i \mid pa_i\right)$$

Sum over all variables not in $\boldsymbol{X} \cup \boldsymbol{Y}$

$$\boxed{P(\mathbf{y} \mid do(\mathbf{x})) = \sum_{\mathbf{v} \setminus (\mathbf{x} \cup \mathbf{y})} \prod_{\{V_i \in \mathbf{V} \setminus \mathbf{X}\}} P\left(v_i \mid pa_i\right)}$$

Adjustment by Direct Parents

Thm 4.2.2. Given a causal diagram G of any Markovian system, the causal quantity $Q = P(\boldsymbol{y} \mid do(\boldsymbol{x}))$ is identifiable whenever $\{\boldsymbol{X}, \boldsymbol{Y}, Pa_{\boldsymbol{x}}\} \subseteq \boldsymbol{V}$, that is, whenever $\boldsymbol{X}$, $\boldsymbol{Y}$, and all the parents of variables $\boldsymbol{X}$ are measured. The expression of Q is then obtained by adjustment for $PA_{\boldsymbol{x}}$, or

$$P(\mathbf{y} \mid do(\mathbf{x})) = \sum_{pa_{\mathbf{x}}} P\left(\mathbf{y} \mid \mathbf{x}, pa_{\mathbf{x}}\right) P\left(pa_{\mathbf{x}}\right)$$

Quiz: derive from previous slide

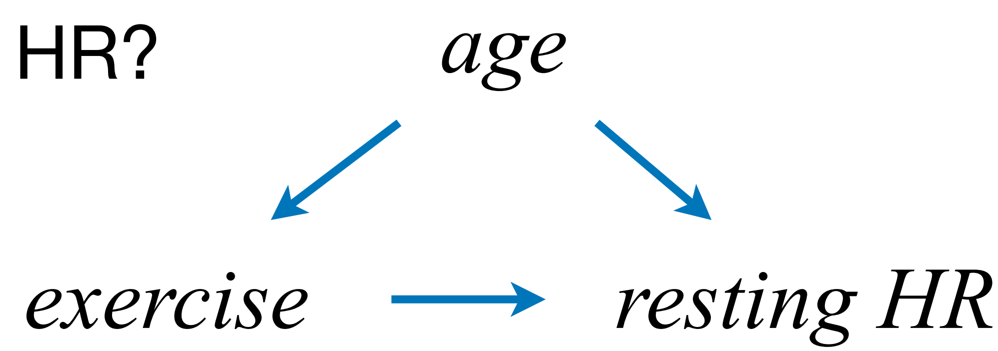
How could adjustment help in real data analysis?

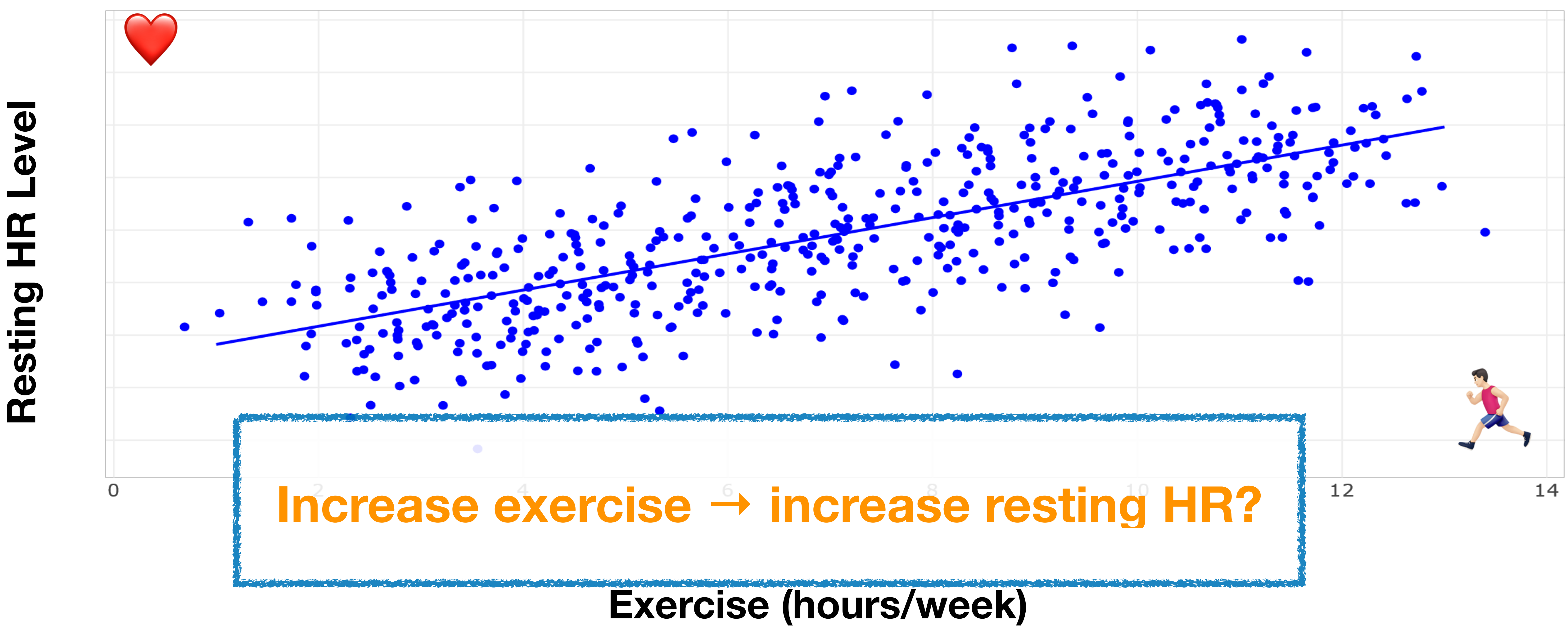
(The Problem of Confounding)

Confounding Bias

What's the causal effect of Exercise on resting HR?

What about $P(\text{resting HR} \mid \text{exercise})$?

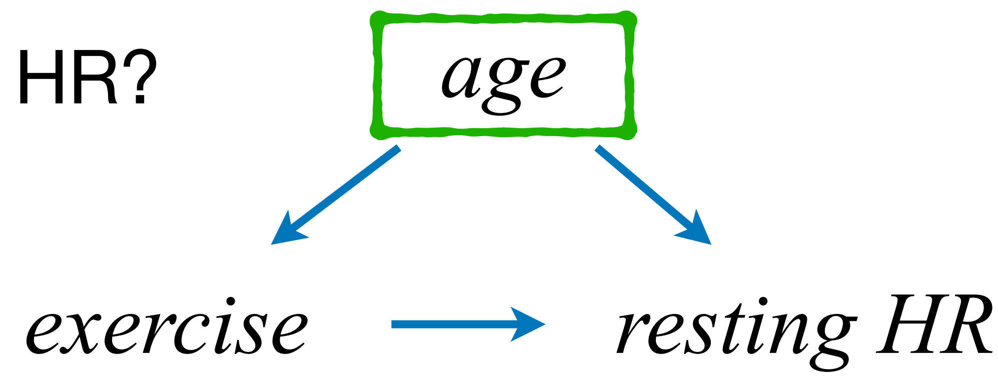


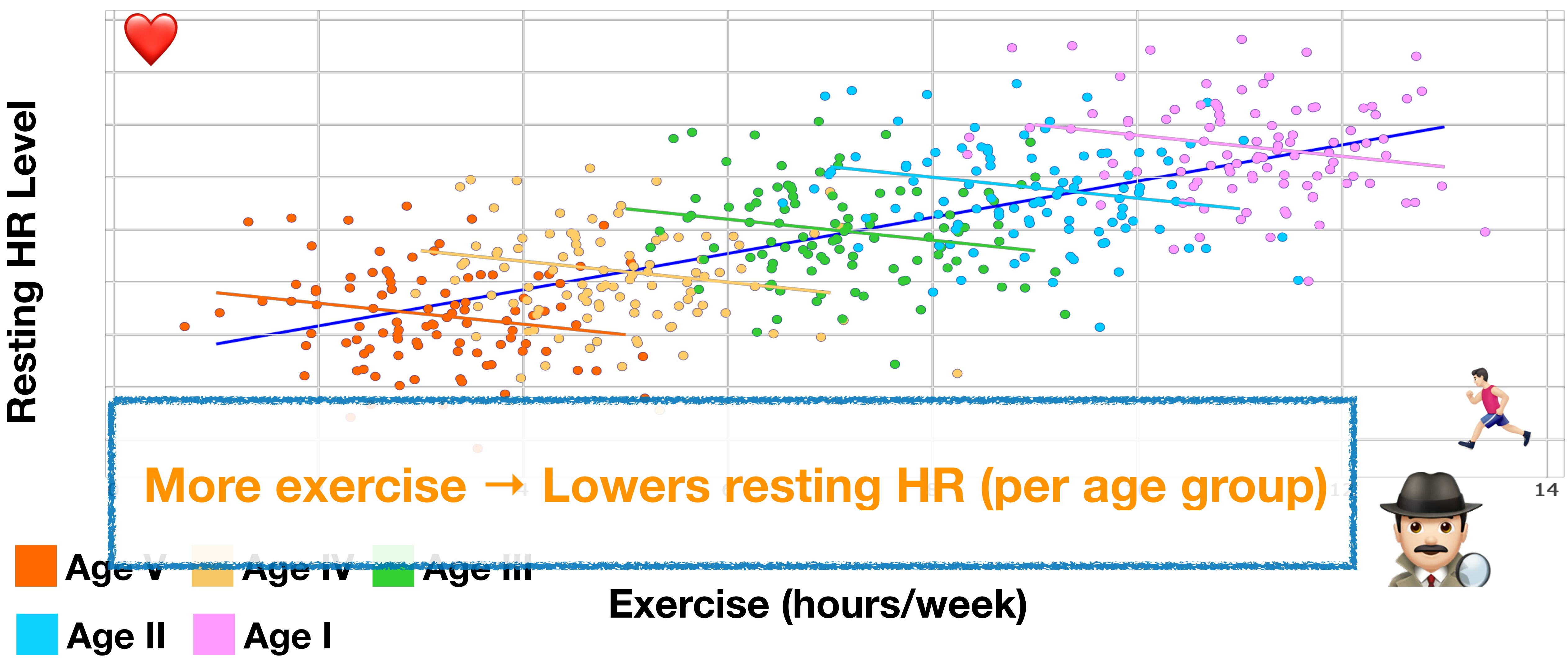


Confounding Bias

What's the causal effect of Exercise on resting HR?

What about *P(resting HR | exercise)* ?



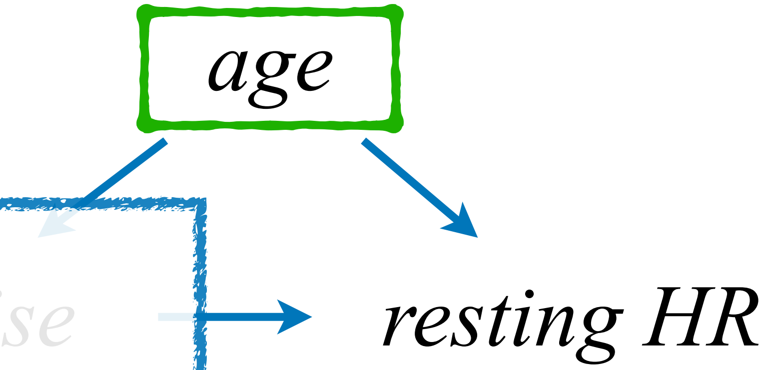


More exercise → Lowers resting HR (per age group)

Confounding Bias

What's the causal effect of Exercise on resting HR?

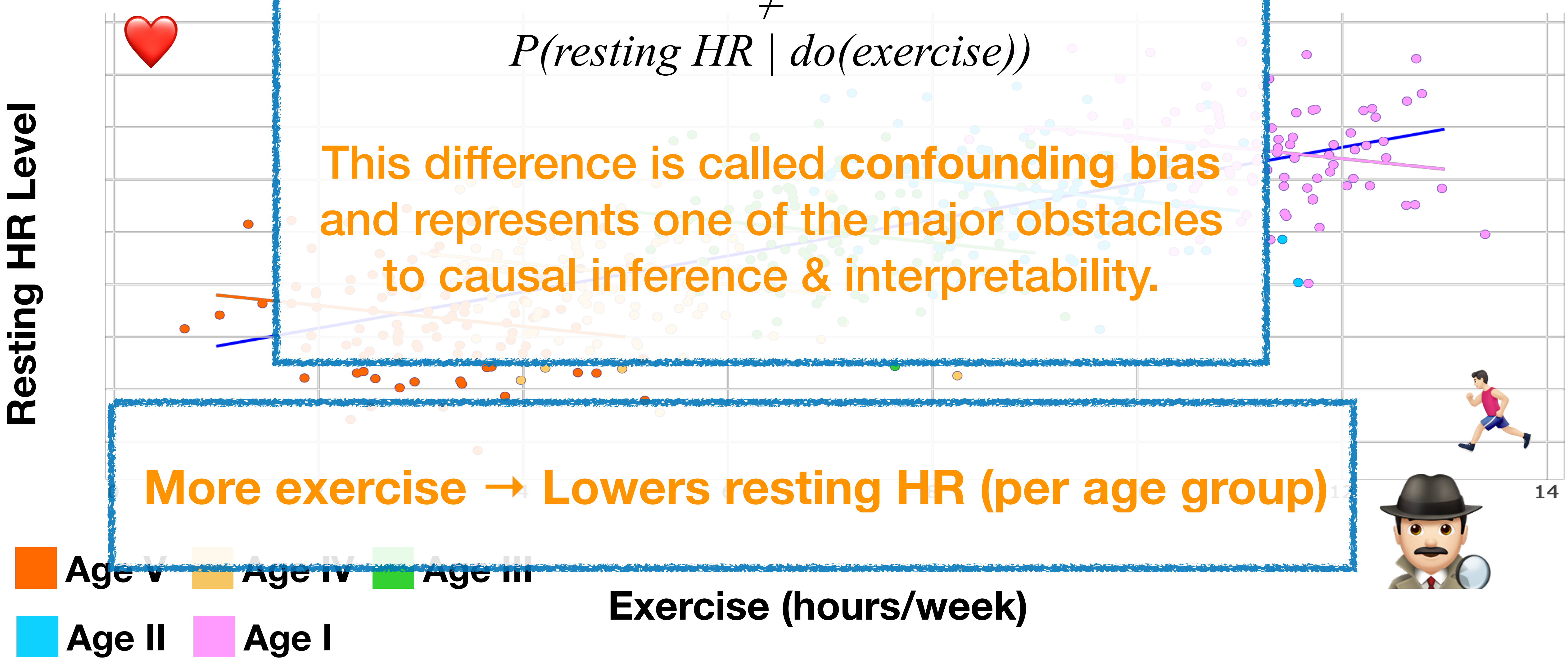
What about *P(resting HR | exercise)* ?



$$P(\textit{resting HR} \mid \textit{exercise}) \neq P(\textit{resting HR} \mid do(\textit{exercise}))$$

This difference is called **confounding bias** and represents one of the major obstacles to causal inference & interpretability.

More exercise → Lowers resting HR (per age group)



If Obesity is latent, is the effect still computable?

Queries:

$$\mathbf{Q_2} = P(dm \mid do(IR = t))$$

$$= \frac{\sum_{ob,htn} P(dm \mid IR = t, htn) P(IR = t) P(htn \mid ob) P(ob)}{P(IR = t)}$$

equal to 1

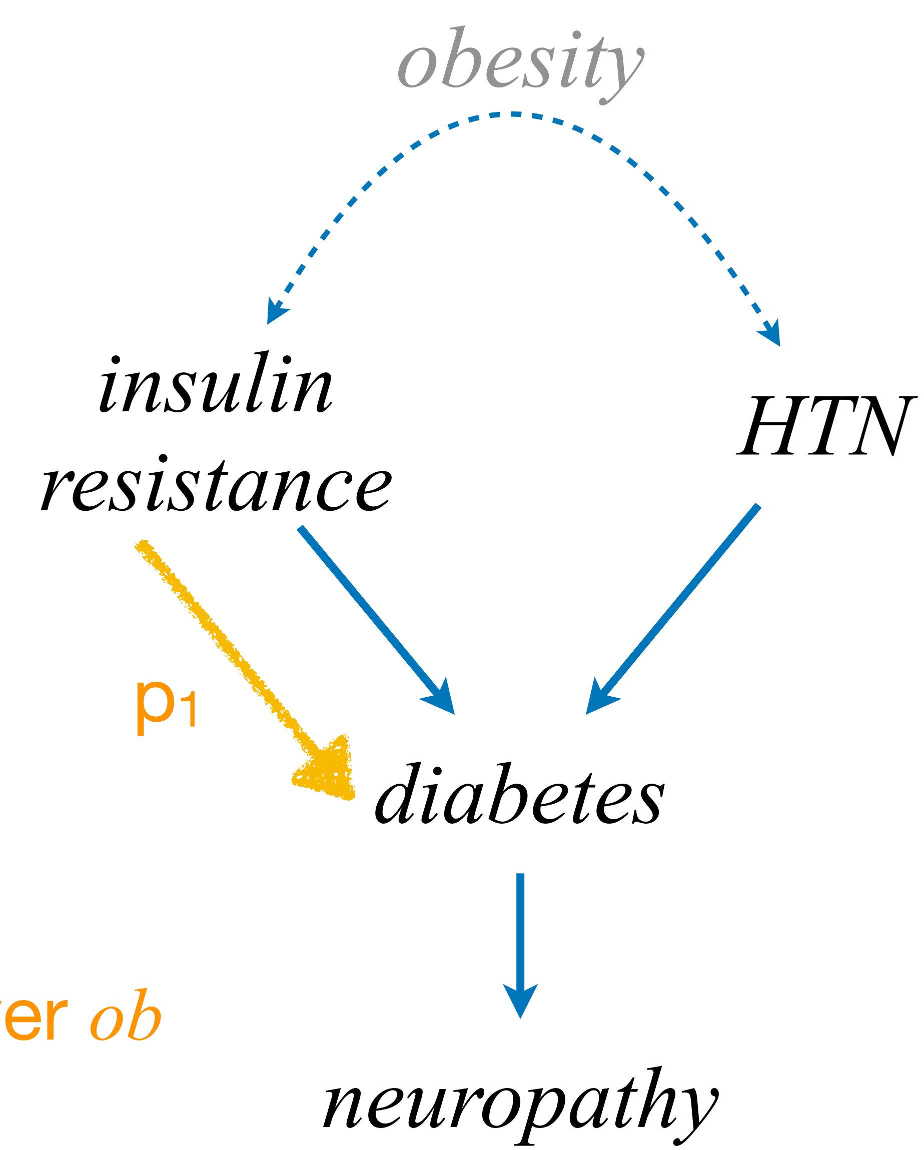
$$= \sum_{ob,htn} P(dm \mid IR = t, htn) P(htn \mid ob) P(ob)$$

$$= \sum_{ob,htn} P(dm \mid IR = t, htn) P(htn, ob)$$ Bayes rule

$$= \sum_{htn} P(dm \mid IR = t, htn) \sum_{ob} P(htn, ob)$$ summing over *ob*

$$\boxed{= \sum_{htn} P(dm \mid IR = t, htn) P(htn)}$$

Adjustment by hypertension!



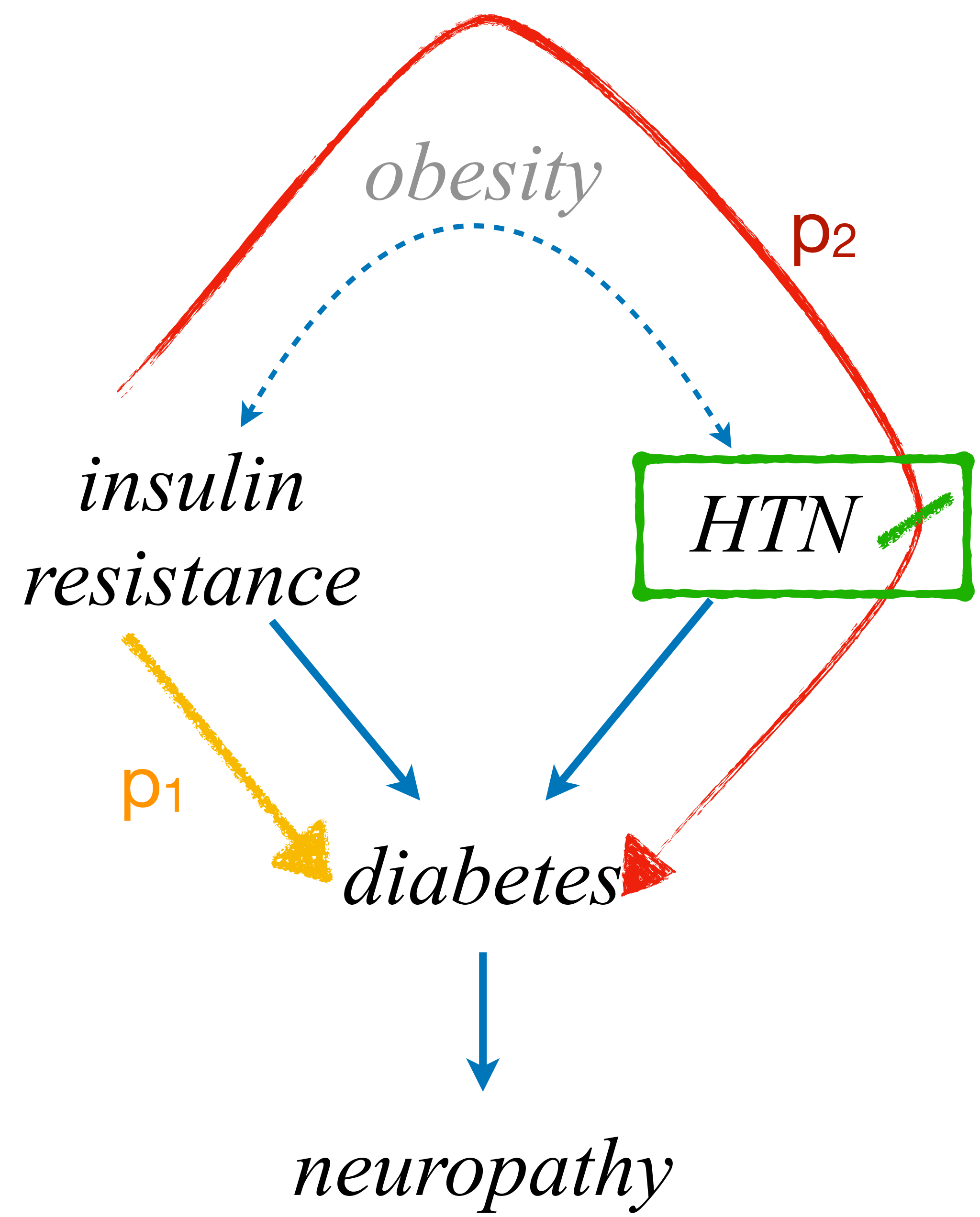
If Obesity is latent, is the effect still computable?

Queries:

$\boldsymbol{Q_2} = P(dm \mid do(IR = t))$

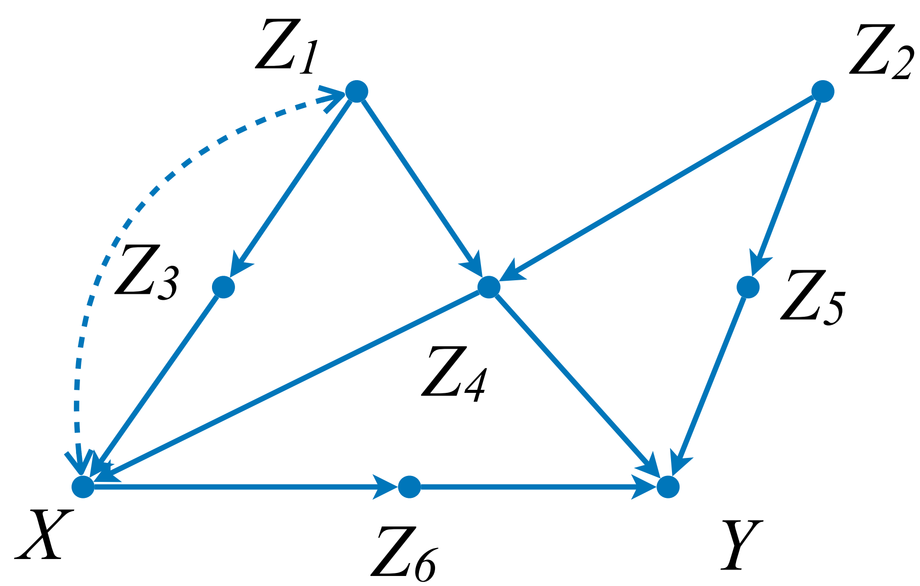
$$\boxed{= \sum_{htn} P(dm \mid IR = t, \underline{htn)P(htn})}$$

By conditioning on hypertension,
- p_2 (the non-causal path) is blocked, and
- p_1 (the causal path) remains unaffected!



Is Confounding Bias removable?

Goal: Find the effect of $\boldsymbol{X}$ on $\boldsymbol{Y}$, $Q = P(\mathbf{y}|do(\mathbf{x}))$, given measurements on variables $Z_1, \ldots, Z_k$,



where some of $\boldsymbol{X}$ parents are unobserved.

How can the target quantity Q be identified if only a subset of the parents is measured?

Answer:
The Back-door Criterion

Definition 4.2.2. (Back-door Criterion)

A set Z satisfies the back-door criterion (bdc) w.r.t. to a pair of variables X, Y in a causal diagram G if:

(i) no node in Z is a descendent of X; and

(ii) Z blocks every path between X and Y that contains an arrow into X.

The Back-door Adjustment

Theorem 4.2.3. (Back-door Adjustment)

If a set Z satisfies the bdc w.r.t the pair X, Y, the effect of X on Y is identifiable and given by:

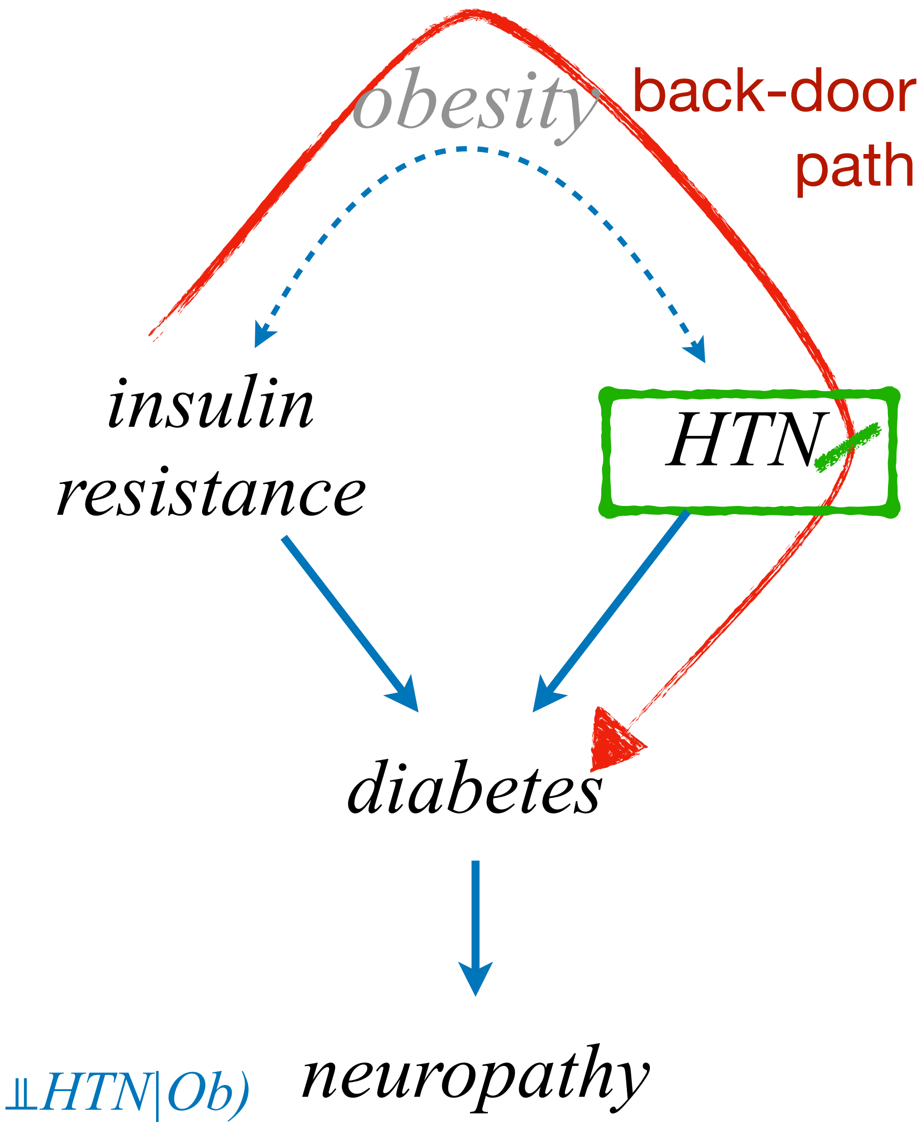
$$P(\mathbf{y} \mid do(\mathbf{x})) = \sum_{\mathbf{z}} P(\mathbf{y} \mid \mathbf{x}, \mathbf{z}) P(\mathbf{z})$$

Back-door sets as substitutes of the direct parents of X

Hypertension satisfies the back-door criterion relative to *ins. resist.* and *diabetes*:

(i) HTN is not a descendant of IR, and

(ii) HTN blocks the only back-door path from IR to *diabetes*.



Adjusting for the direct parents of IR, we have:

$$\begin{aligned}
P(dm \mid do(ir)) &= \sum_{ob} P(dm \mid ir, ob) P(ob) \\
&= \sum_{ob,htn} P(dm \mid ir, ob, htn) P(htn \mid ir, ob) P(ob) \\
&= \sum_{ob,htn} P(dm \mid ir, htn) P(htn \mid ob) P(ob) \\
&= \sum_{htn} P(dm \mid ir, htn) \sum_{ob} P(htn, ob) = \boxed{\sum_{htn} P(dm \mid ir, htn) P(htn)}
\end{aligned}$$

$(IR \perp\!\!\!\perp HTN \mid Ob)$

$(Ob \perp\!\!\!\perp DM \mid HTN, IR)$

Adjustment by *Hypertension*

Adjustment by Direct Parents → Back-door Adjustment

More Generally:

(i) no node in Z is a descendent of X; and $\Longrightarrow$ $(Z \perp\!\!\!\perp X \mid Pa_x)$

(ii) Z blocks every path between X and Y that contains an arrow into X. $\Longrightarrow$ $(Y \perp\!\!\!\perp Pa_x \mid Z, X)$

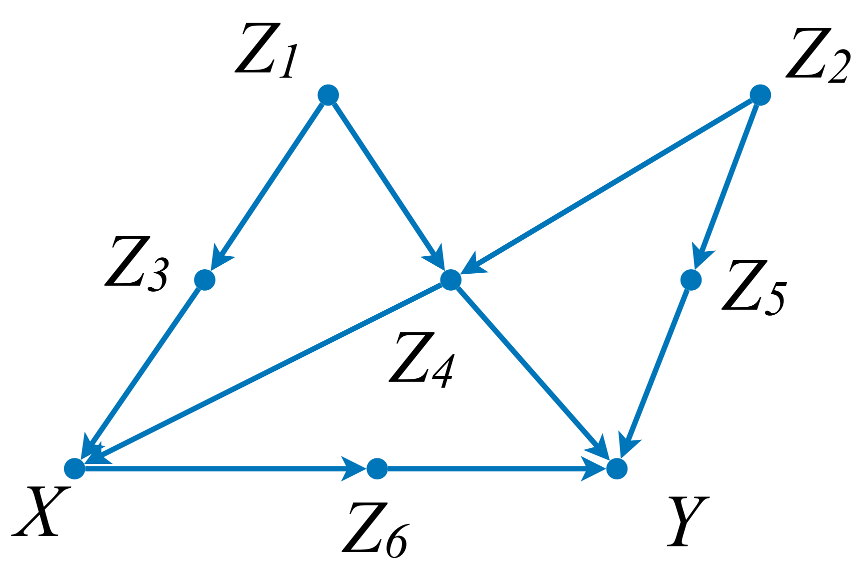
Then:

$$
\begin{aligned}
P(\mathbf{y} \mid do(\mathbf{x})) &= \sum_{pa_{\mathbf{x}}} P(\mathbf{y} \mid \mathbf{x}, pa_{\mathbf{x}}) P(pa_{\mathbf{x}}) \\
&= \sum_{\mathbf{z}, pa_{\mathbf{x}}} P(\mathbf{y} \mid \mathbf{x}, pa_{\mathbf{x}}, \mathbf{z}) P(\mathbf{z} \mid \mathbf{x}, pa_{\mathbf{x}}) P(pa_{\mathbf{x}}) \\
&= \sum_{\mathbf{z}, pa_{\mathbf{x}}} P(\mathbf{y} \mid \mathbf{x}, \mathbf{z}) P(\mathbf{z} \mid pa_{\mathbf{x}}) P(pa_{\mathbf{x}}) \\
&= \sum_{\mathbf{z}} P(\mathbf{y} \mid \mathbf{x}, \mathbf{z}) \sum_{pa_{\mathbf{x}}} P(\mathbf{z}, pa_{\mathbf{x}}) = \boxed{\sum_{\mathbf{z}} P(\mathbf{y} \mid \mathbf{x}, \mathbf{z}) P(\mathbf{z})}
\end{aligned}
$$

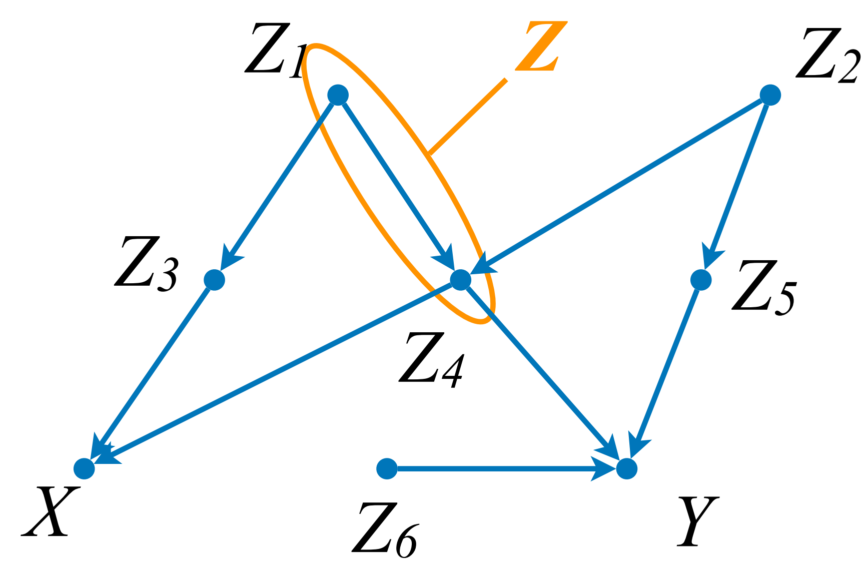
Adjustment by Z is equivalent to adjustment by direct parents whenever Z is bd-admissible!

How do we find these bd-sets? Graphical Condition

$P(y \mid do(x))$ is identifiable if there is a set $\boldsymbol{Z}$ that d-separates $\boldsymbol{X}$ from $\boldsymbol{Y}$ in $G_{\underline{x}}$



G

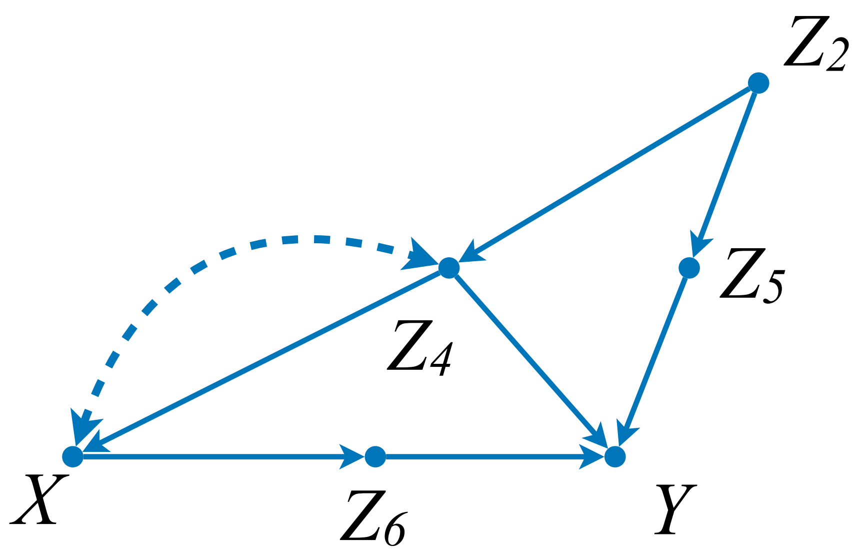


$G_{\underline{x}}$

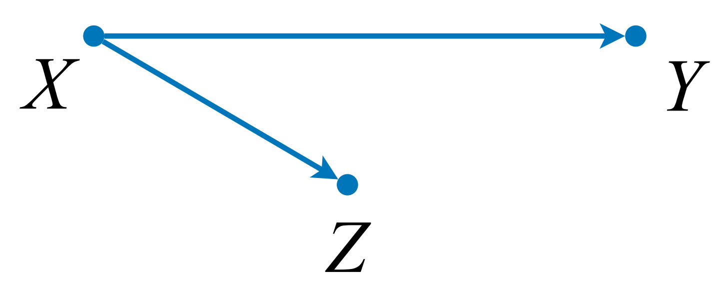
$$P(y \mid do(x)) = \sum_{z_1, z_4} P(y \mid x, z_1, z_4) P(z_1, z_4)$$

Further Back-door Examples

Are there admissible back-door sets (relative to $\boldsymbol{X}, \boldsymbol{Y}$) for the following graphs?



$\boldsymbol{Z} = \{Z_4, Z_2\}, \{Z_4, Z_5\},$
$\{Z_4, Z_2, Z_5\}$

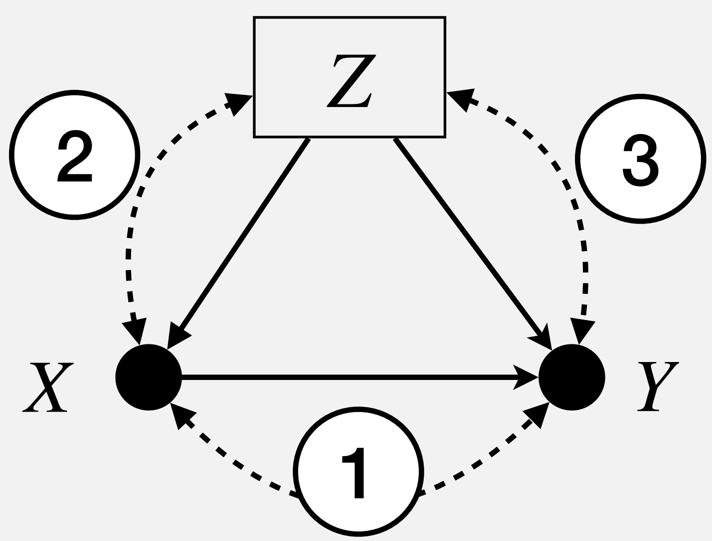


$\boldsymbol{Z} = \varnothing$

Most Common Back-door Setting: CG(3) Graphical Model

- What is the simplest setting in which back-door helps?
  Approach: put all Z variables into a single block!

Construct graph over three blocks



start with no assumptions

CG(3) model

Elicit Causal Knowledge Stepwise

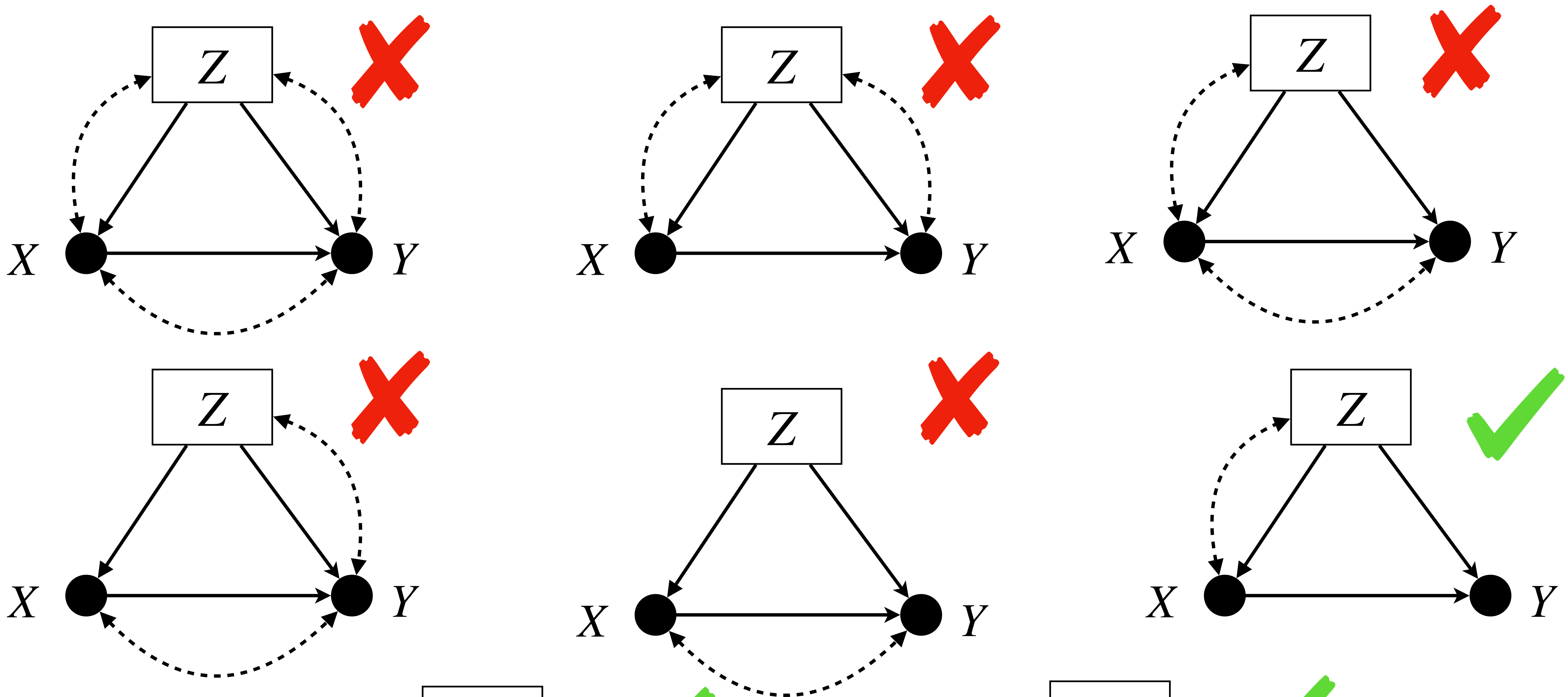
1 is there confounding X - Y?

2 is there confounding X - Z?

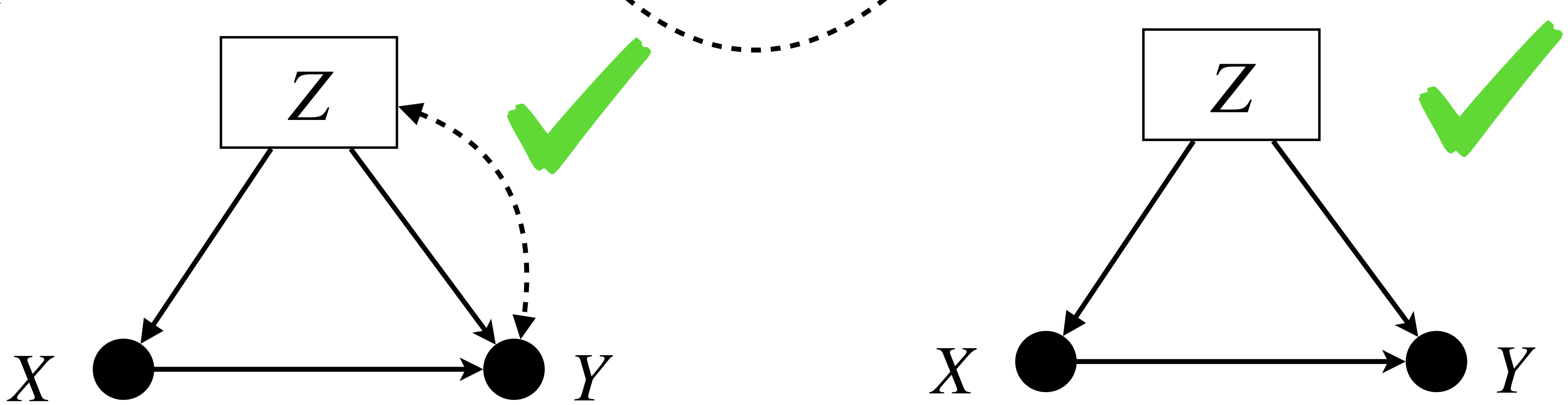
3 is there confounding Y - Z?

Causal Assumptions encoded in edge removals!

Assess Back-door: 8 possible cases



If Back-Door holds for Z, can use it for adjustment!

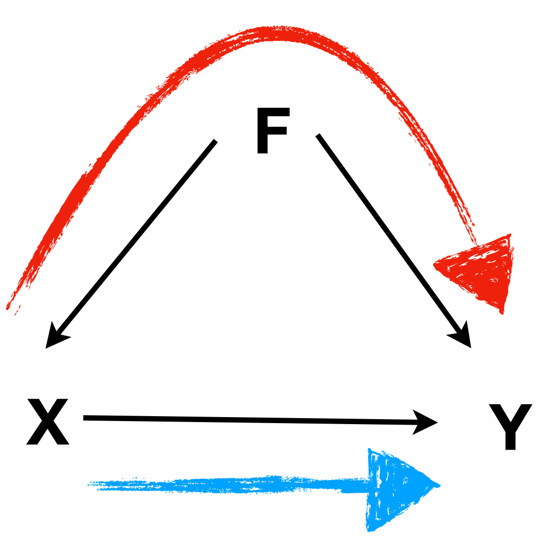


Back to Sex-Stratification vs. Weight-Loss Stratification: Which table to consult?

$(\neg F)$	HbA1c low (Y)	HbA1c high $(\neg Y)$		Sucess Rate
drug (X)	18	12	30	60%
no-drug $(\neg X)$	7	3	10	70%
	25	15	40	

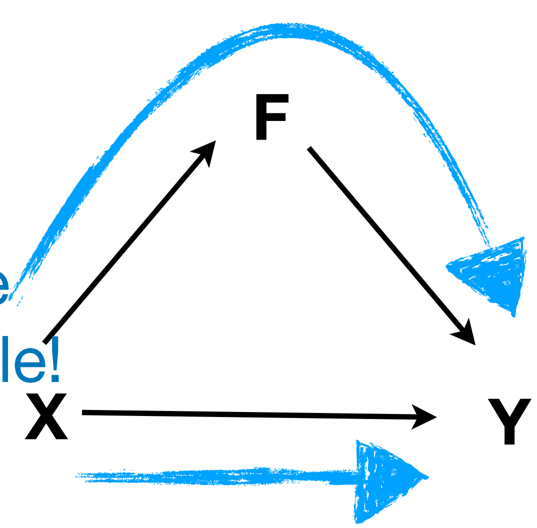
(F)	HbA1c low (Y)	HbA1c high $(\neg Y)$		Sucess Rate
drug (X)	2	8	10	20%
no-drug $(\neg X)$	9	21	30	30%
	11	29	40	

Sex-Stratified



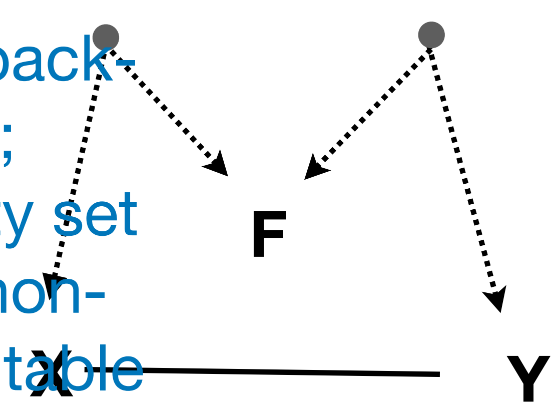
Weight Loss-Stratified

F is back-door; consult the stratified table!



Another Story?

F is not back-door; but empty set is! use non-stratified table



FFT