Causal Inference for Health Data

(STATS C160/C260 — Winter 2026)

Lecture 4: Identification of Causal Effects

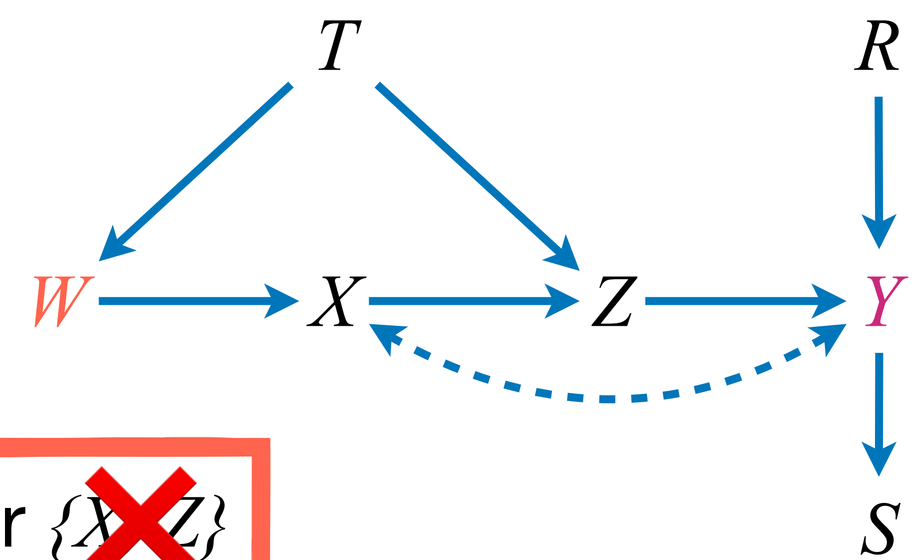
Drago Plečko

UCLA College | Physical Sciences
Statistics & Data Science

Causal Inference FOR HEALTH DATA

Recap:
Graph Separation (d-Separation)

Is there a set A such that the separation statement $(W \perp\!\!\!\perp Y \mid A)$ holds?

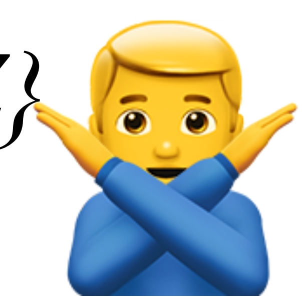


Path 1: $W \longrightarrow X \longrightarrow Z \longrightarrow Y$ — ~~$\{X\}$~~ or $\{Z\}$ or ~~$\{X, Z\}$~~

Path 2: $W \longleftarrow T \longrightarrow Z \longrightarrow Y$ — $\{T\}$ or $\{Z\}$ or $\{T, Z\}$

Path 3: $W \longrightarrow X \longleftrightarrow Y$ — not X

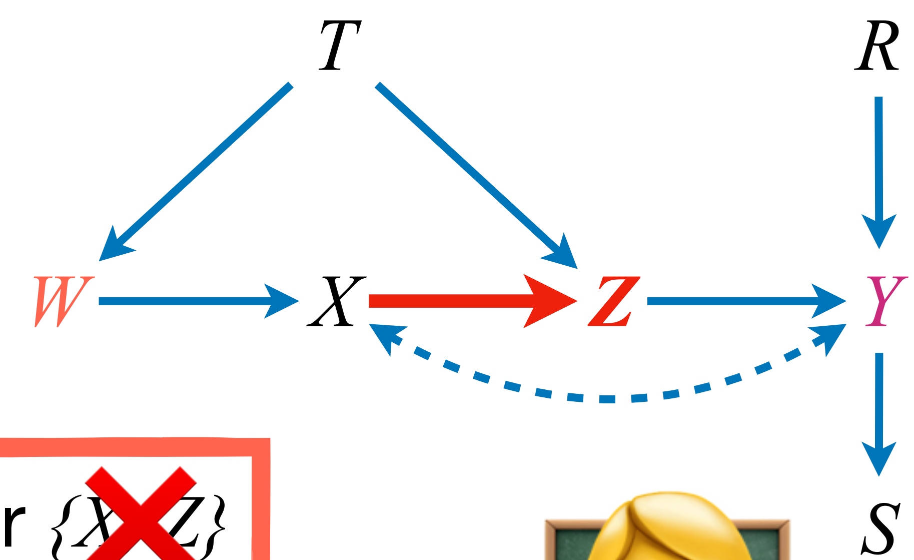
Does $A = \{T, Z\}$ suffice?



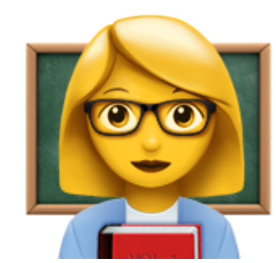
Path 4: $W \longleftarrow T \longrightarrow Z \longleftarrow X \longleftrightarrow Y$ — $\{T\}$ or ~~$\{X\}$~~ or ~~$\{T, X\}$~~ or $\{T, Z\}$ or ~~$\{T, X, Z\}$~~

Recap: Graph Separation (d-Separation)

Is there a set A such that the separation statement $(W \perp\!\!\!\perp Y \mid A)$ holds?



Path 1: $W \longrightarrow X \longrightarrow Z \longrightarrow Y$ ~~$\{X\}$~~ or ~~$\{Z\}$~~ or ~~$\{X,Z\}$~~



Path 2: $W \longleftarrow T \longrightarrow Z \longrightarrow Y$ $\{T\}$ or ~~$\{Z\}$~~ or ~~$\{T,Z\}$~~

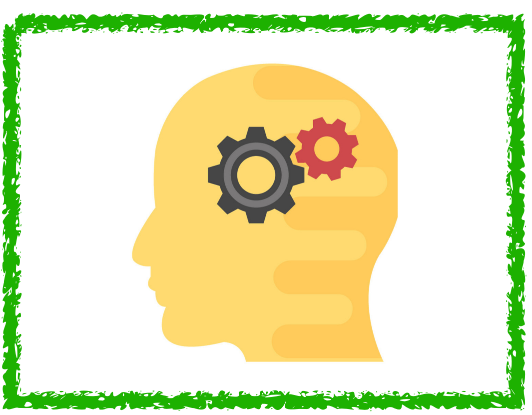
No such A!

Don't forget the descendants of the colliders!

Path 3: $W \longrightarrow X \longleftrightarrow Y$, $X \longrightarrow Z$ not X | not Z

Path 4: $W \longleftarrow T \longrightarrow Z \longleftarrow X \longleftrightarrow Y$ $\{T\}$ or ~~$\{X\}$~~ or ~~$\{T,X\}$~~ or ~~$\{T,Z\}$~~ or ~~$\{T,X,Z\}$~~

3. SCM → Pearl's Causal Hierarchy



SCM M

Level (Symbol)	Typical Activity
1 👀 Association $P(y \mid x)$	Seeing Observational Data
2 💪 Intervention $P(y \mid do(x), c)$	Doing Experimental Data
3 🧠 Counterfactual $P(y_x \mid x', y')$	Imagining, Retrospection

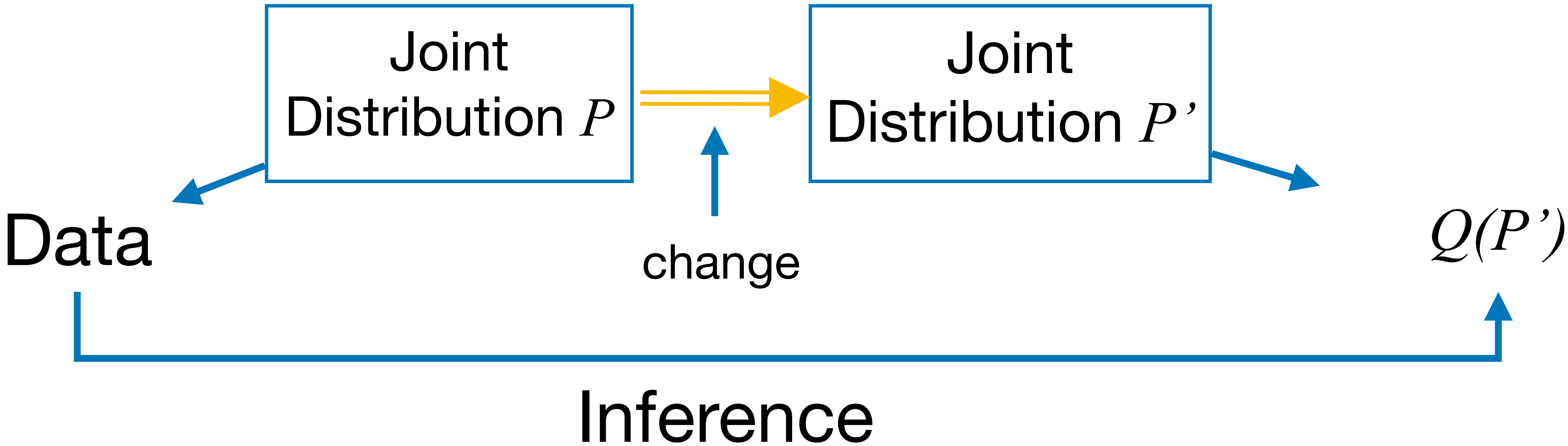
2nd Layer of the Causal Hierarchy

Causal Effects

(What if I do $X{=}x$?)

CBSNEWS
Science News
from universities, journals, and other research organizations
One Drink Of Red W
ing To Circulation, But Two
Drinks Are Stressf
Feb. 13, 2008 — One dr
alcohol slightly
but the positive
disappear with
Peter Munk Ca
Hospital.
INSIDE THE EM
OF BOYS
Home
News
Current Issue
Why D
One of the
the consis
By John C
Correc
Dr. Sa
August 15t
08:00 PM E
THE CON
Gra
Re
By RON
PRINT
In Texas it is illegal to take
41 Comments
Shares
The NEW ENGLAND JOURNAL of MEDICINE
HOME
ARTICLES & MULTIMEDIA
ISSUES
SPECIALTIES & TOPICS
FOR AUTHORS
CME
ORIGINAL ARTICLE
Association of Nut Consumption with Total and Cause-Specific Mortality
Ying Bao, M.D., Sc.D., Jiali Han, Ph.D., Frank B. Hu, M.D., Ph.D., Edward L. Giovannucci, M.D., Sc.D., Meir J. Stampfer, M.D., Dr.P.H., Walter C. Willett, M.D., Dr.P.H., and Charles S. Fuchs, M.D., M.P.H.
N Engl J Med 2013; 369:2001-2011 | November 21, 2013 | DOI: 10.1056/NEJMoa1307352
Abstract
Article
References
BACKGROUND
Increased nut consumption has been associated with a reduced risk of major chronic diseases, including cardiovascular disease and type 2 diabetes mellitus. However, the association between nut consumption and mortality remains unclear.
Full Text of Background...
METHODS
We examined the association between nut consumption and
Share:
MEDIA IN THIS ARTICLE
Video
Nuts
Nuts and D

Causal Inference — Connecting Different Worlds



What happens when P changes?

e.g., Infer whether less people would get cancer
if we ban smoking.

$Q = P(Cancer = true \mid do(Smoking = no))$ Not an aspect of P.

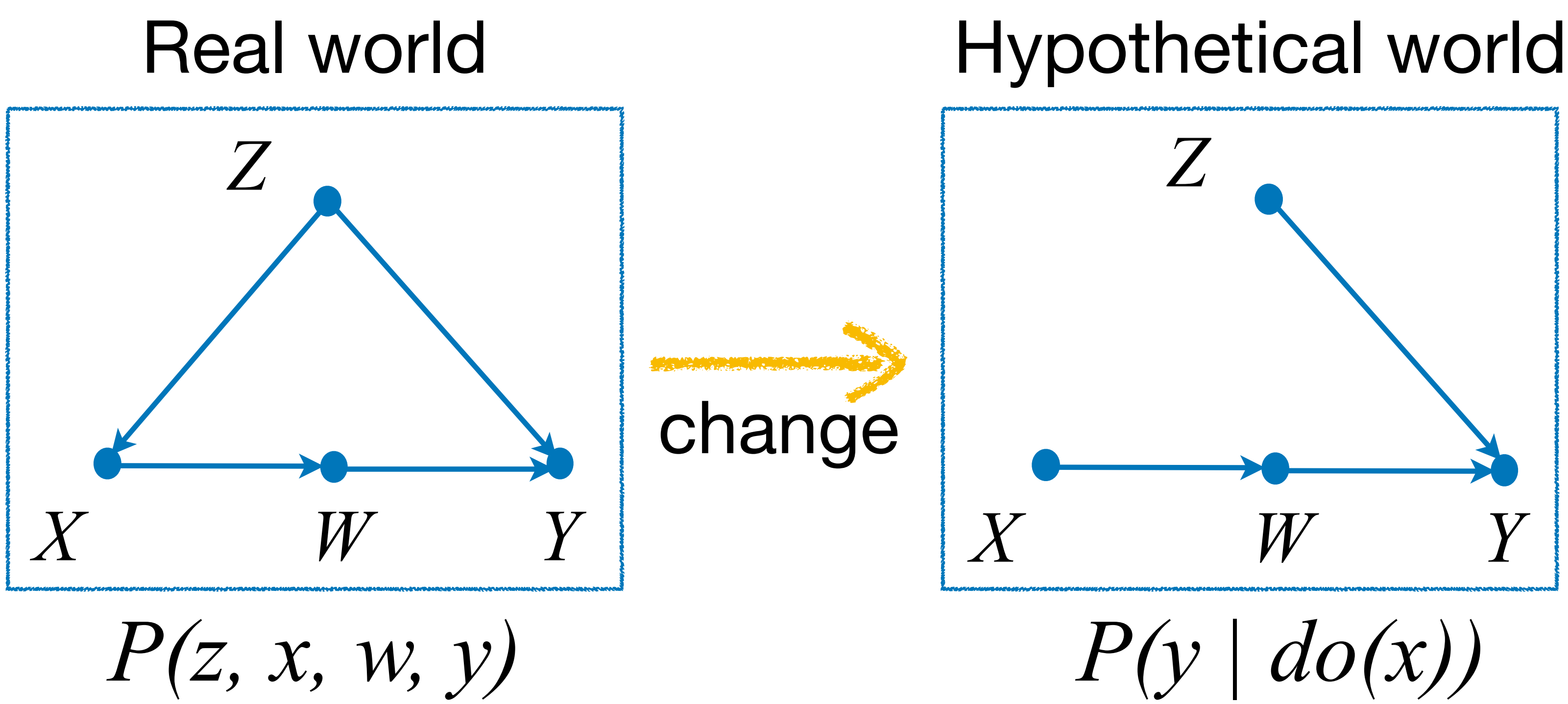
Observation 1:

The observational distribution P alone tells us nothing about change; it just describes static conditions of a population we are passively observing.

Observation 2:

We need to be able to represent "change", or how the population reacts when it undergoes a specific intervention.

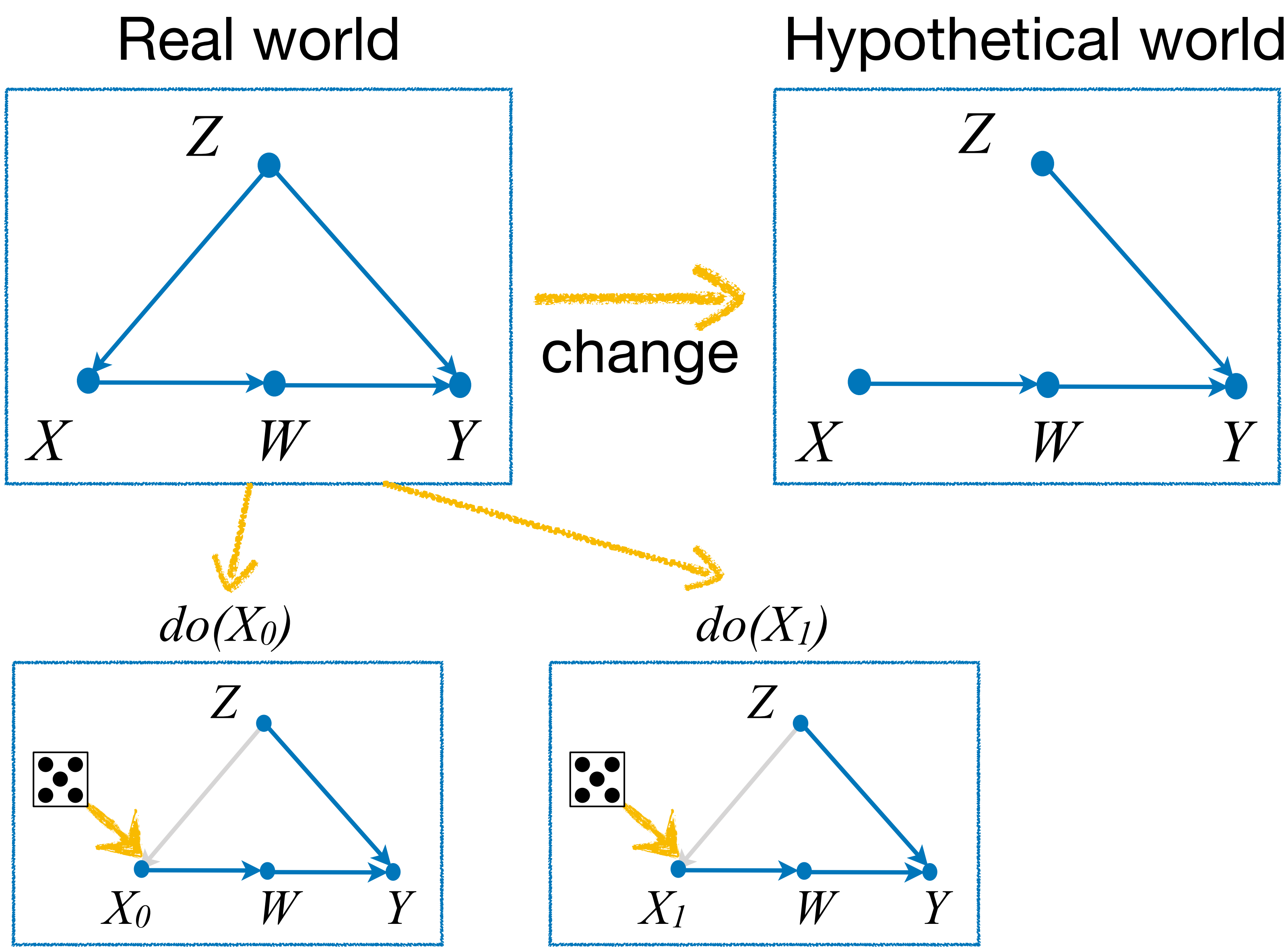
The Challenge of Causal Inference



Z : age, sex
X : action
W : mediator
Y : outcome

- Goal: how much Y changes with X if we vary X between two different constants free from the influence of Z (e.g., treatment vs. no treatment).
- These variations are called causal effects!

Method for Computing Causal Effects: Randomized Experiments



Z : age, sex

X : action

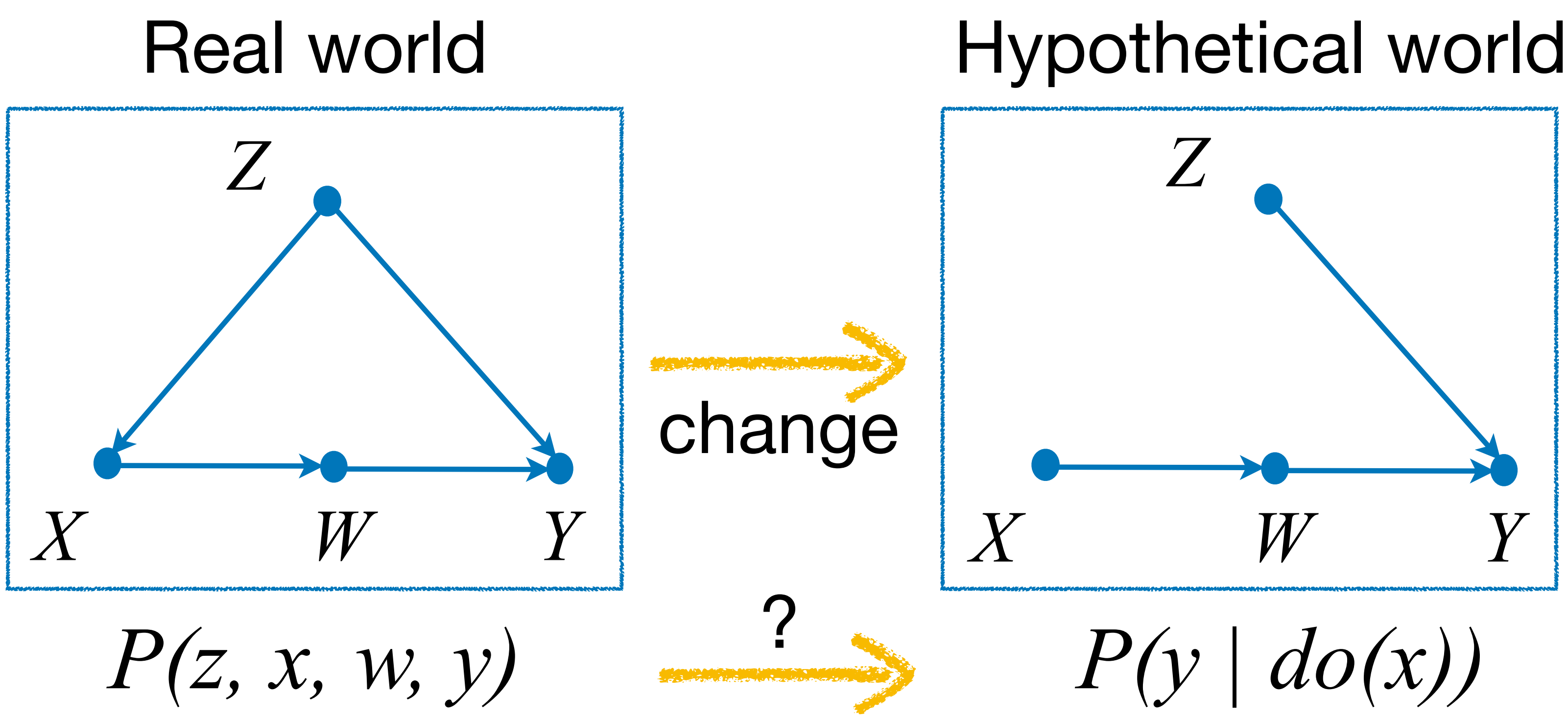
W : mediator

Y : outcome

Randomization:

$P(y \mid do(X_0))$ $P(y \mid do(X_1))$

Computing Causal Effects (L_2) from Observational Data (L_1)



Z : age, sex

X : action

W : mediator

Y : outcome

Questions:

* What is the relationship between $P(z, x, w, y)$ and $P(y \mid do(x))$?
* Is $P(y \mid do(x)) = P(y \mid x)$?

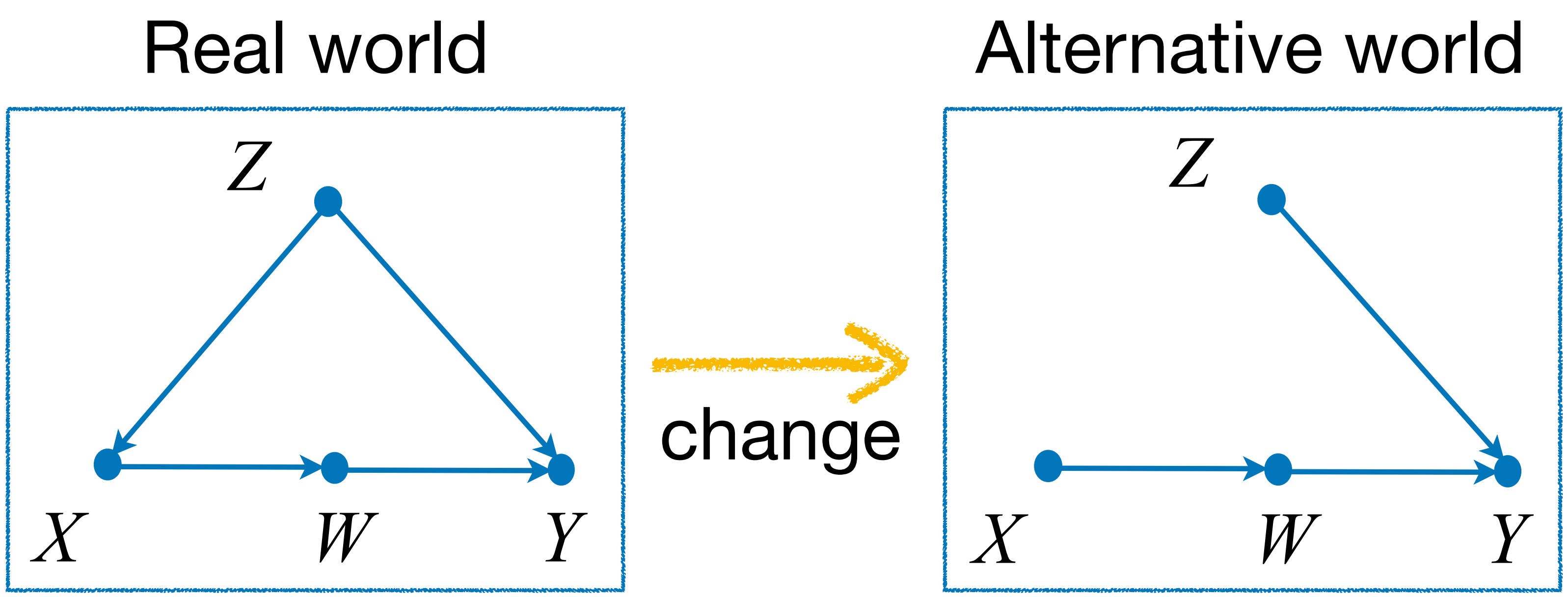
Causal Effects (formal)

Causal Effect (Def. 4.1.1):

Given two disjoint sets of variables, X and Y, the causal effect of X on Y, denoted as $P(y \mid do(x))$, is a function from X to the space of probability distributions of Y.

For each realization x of X, $P(y \mid do(x))$ gives the probability $Y = y$ induced by deleting from the model all equations corresponding to variables in X and substituting $X = x$ in the remaining equations.

Computing Causal Effects from Observational Data



Z : age, sex

X : action

W : mediator

Y : outcome

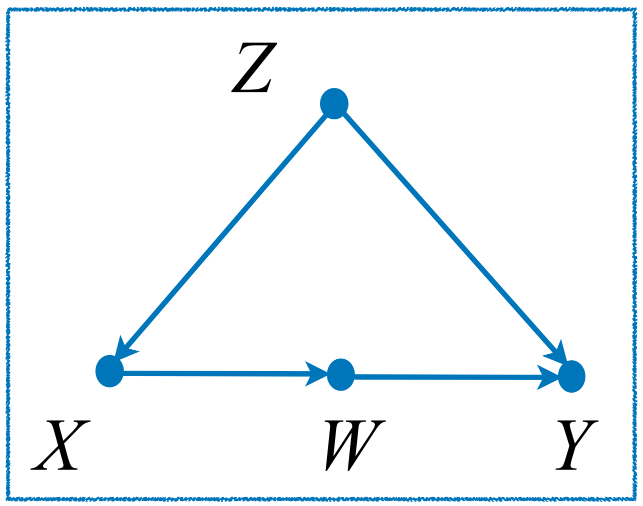
$$M = \begin{cases} Z \leftarrow f_Z(u_z) \\ X \leftarrow f_X(z, u_x) \\ W \leftarrow f_W(x, u_w) \\ Y \leftarrow f_Y(w, z, u_y) \end{cases}$$

do(X=x)

$$M_x = \begin{cases} Z \leftarrow f_Z(u_z) \\ \cancel{X \leftarrow f_X(z, u_x)} \quad X = x \\ W \leftarrow f_W(x, u_w) \\ Y \leftarrow f_Y(w, z, u_y) \end{cases}$$

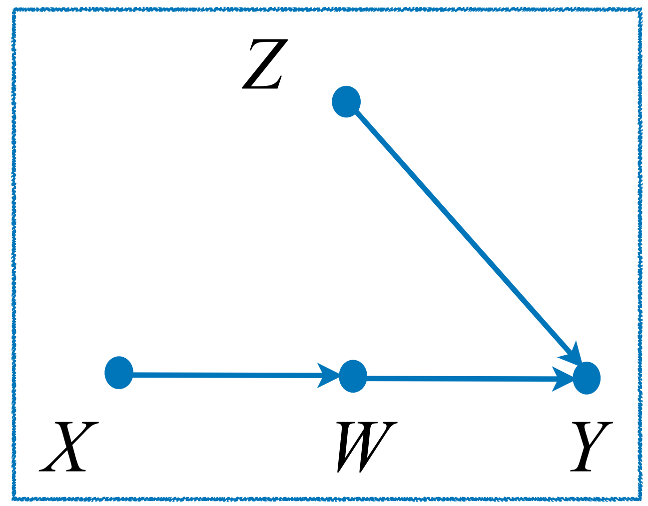
Computing Causal Effects from Observational Data

Real world



change →

Alternative world



Z : age, sex

X : action

W : mediator

Y : outcome

$P(v) =$

$P(z) \times$
$P(x \mid z) \times$
$P(w \mid x) \times$
$P(y \mid w, z)$

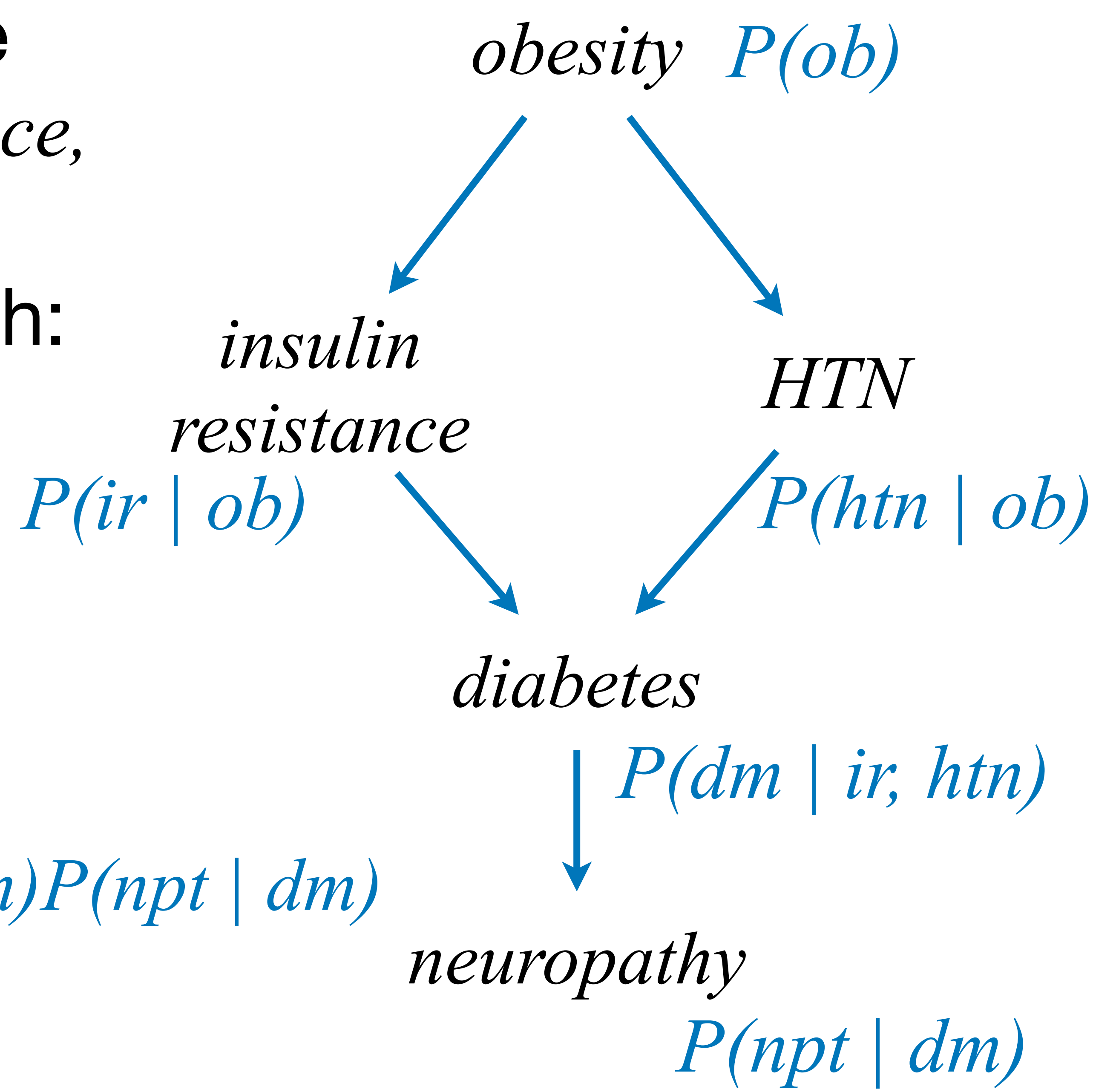
→ $do(X{=}x)$

$P_x(v) =$

$P(z) \times$
~~$P(x)$~~ $\times$ equal to 1 in M_x
$P(w \mid x) \times$
$P(y \mid w, z)$

Computing Causal Effects from Observational Data

Consider a distribution over the variables: *obesity, insulin resistance, hypertension, diabetes*, and *neuropathy*; and the causal graph:



This distribution decomposes as

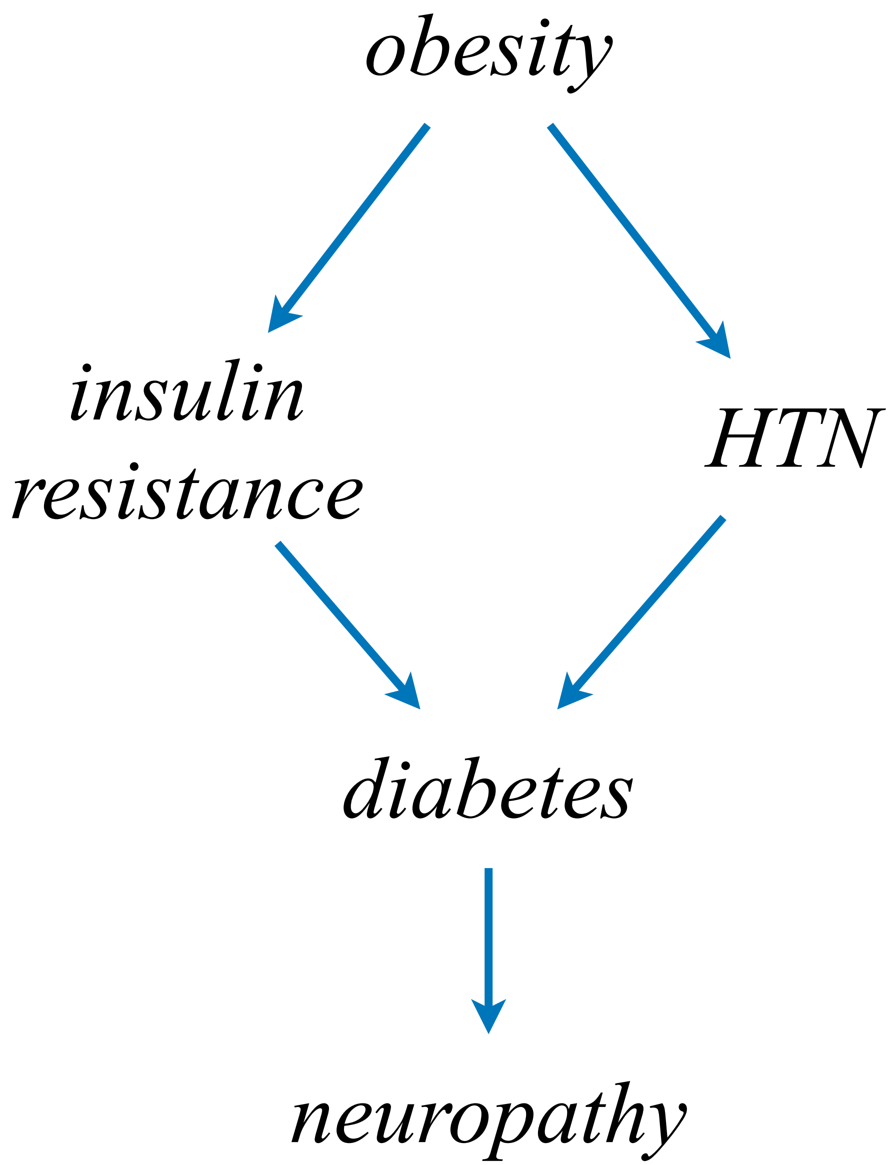
$$P(\mathbf{v}) = P(ob)P(ir \mid ob)P(htn \mid ob)P(dm \mid ir, htn)P(npt \mid dm)$$

Computing Causal Effects from Observational Data

Queries:

$Q_1 = P(dm \mid IR = true)$

$Q_2 = P(dm \mid do(IR = true))$

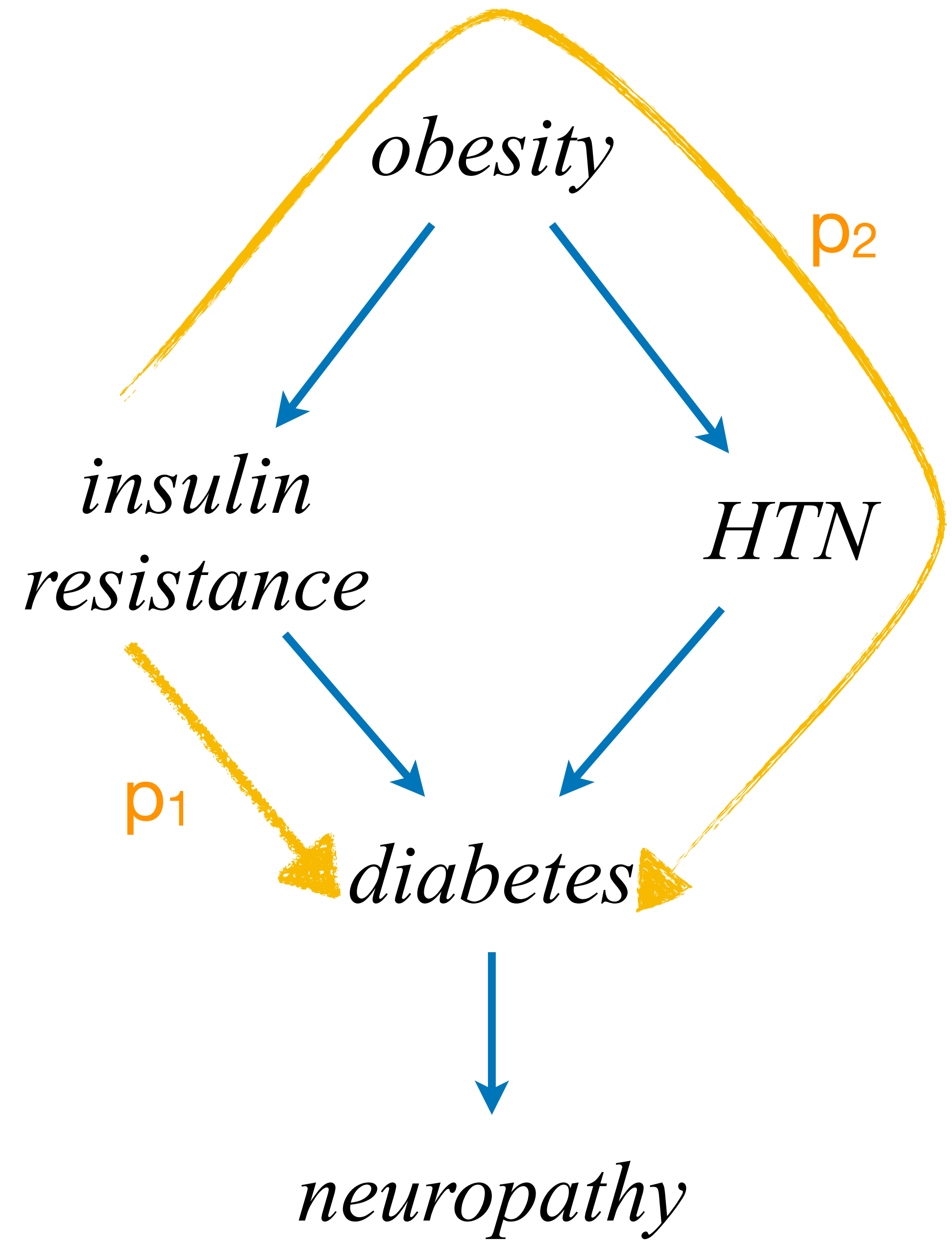


Computing Causal Effects from Observational Data

Queries:

$Q_1 = P(dm \mid IR = true)$
$= \text{“}P(p_1) + P(p_2)\text{”}$

$Q_2 = P(dm \mid do(IR = true))$



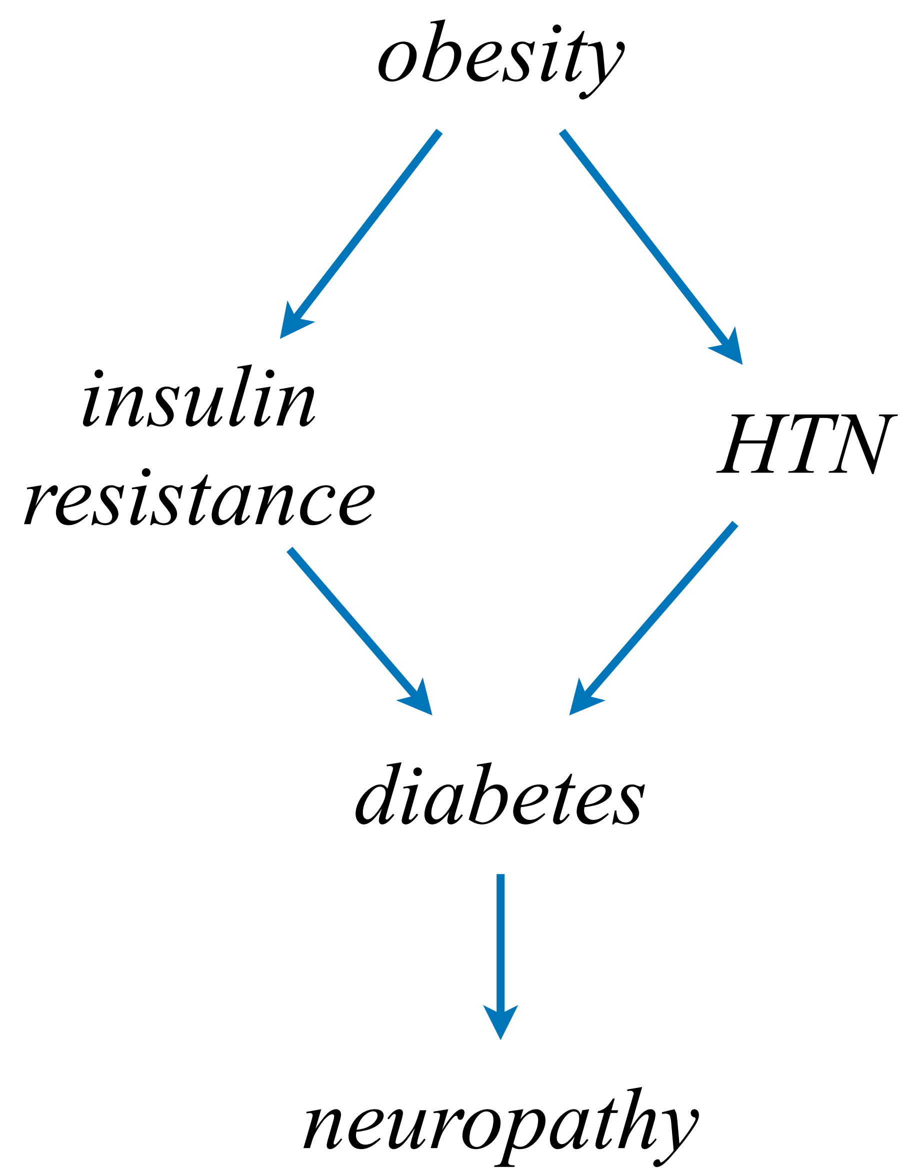
Computing Causal Effects from Observational Data

Queries:

$Q_1 = P(dm \mid IR = true)$

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

$Q_2 = P(dm \mid do(IR = true))$



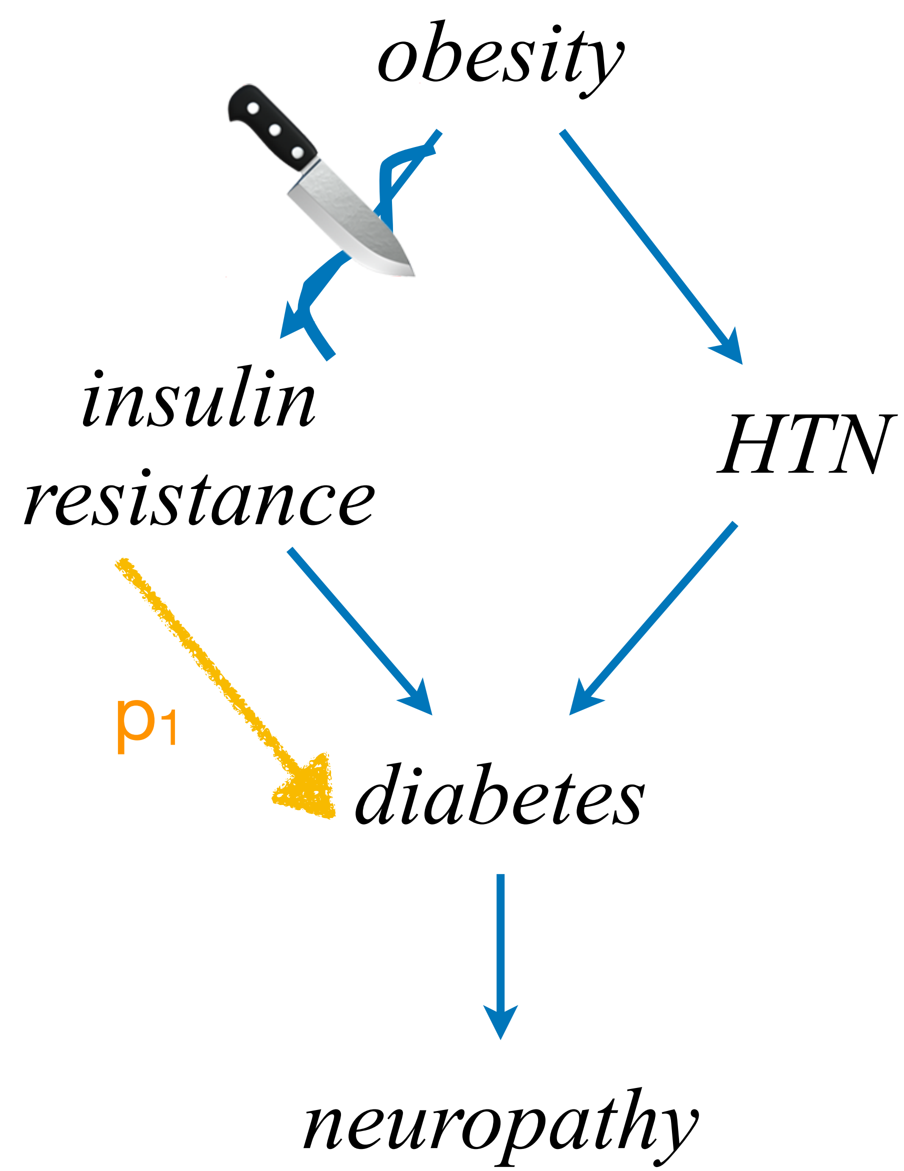
Computing Causal Effects from Observational Data

Queries:

$Q_1 = P(dm \mid IR = true)$

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

$Q_2 = P(dm \mid do(IR = true))$



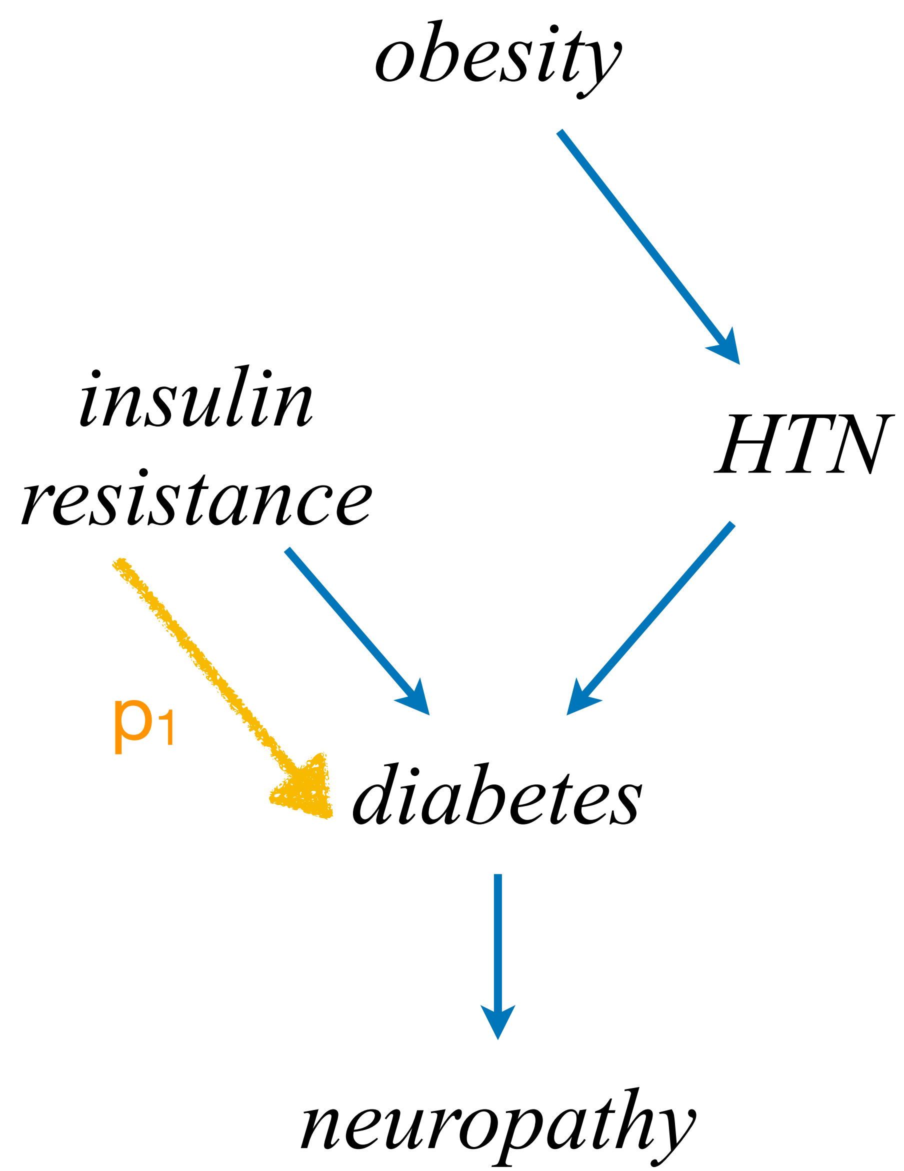
Computing Causal Effects from Observational Data

Queries:

$Q_1 = P(dm \mid IR = true)$

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

$Q_2 = P(dm \mid do(IR = true))$
$= \text{“}P(p_1)\text{”}$



Computing Causal Effects from Observational Data

Queries:

$$Q_1 = P(dm \mid IR = true)$$

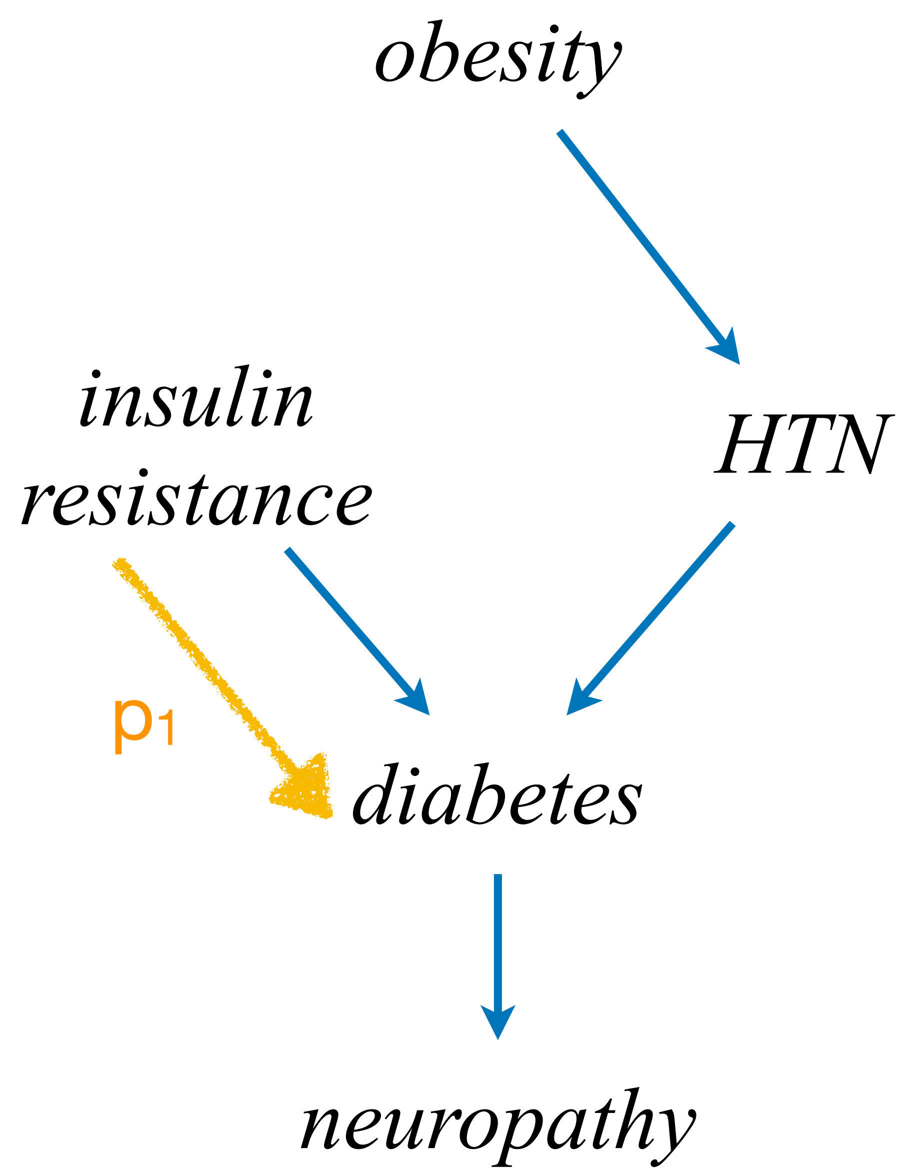
$$= \frac{\sum_{ob,htn} P(dm \mid IR = t, htn) P(IR = t \mid ob) P(htn \mid ob) P(ob)}{\sum_{ob} P(IR = t \mid ob) P(ob)}$$

$$Q_2 = P(dm \mid do(IR = true))$$

$$= \frac{\sum_{ob,htn} P'(dm \mid IR = t, htn) P'(IR = t) P'(htn \mid ob) P'(ob)}{\sum_{ob} P'(IR = t) P'(ob)}$$

equal to 1

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



Truncated Factorization Product (Operationalizing Interventions)

Corollary (Truncated Factorization (Markovian case); Thm. 4.2.1):

The distribution generated by an intervention $do(\mathbf{X}=\mathbf{x})$ (in a Markovian model M) is given by the truncated factorization:

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

product ranges over all $V_i \in \mathbf{V}$ that are not in $\mathbf{X}$

for the observational distribution, the product ranges over all $V_i \in \mathbf{V}$

Truncated Factorization Formula

The truncated product,

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

can be rewritten as:

$$P(\mathbf{v} \mid do(\mathbf{x})) = \frac{P(\mathbf{v})}{P\left(\mathbf{x} \mid pa_{\mathbf{x}}\right)} \Bigg|_{\mathbf{X}=\mathbf{x}}$$

Also equivalent to:

$$P(\mathbf{v} \mid do(\mathbf{x})) = P\left(\mathbf{v} \mid \mathbf{x}, pa_{\mathbf{x}}\right) P\left(pa_{\mathbf{x}}\right) \Big|_{\mathbf{X}=\mathbf{x}}$$

The transformation between the observation and interventional distributions can be seen as a re-weighing process.

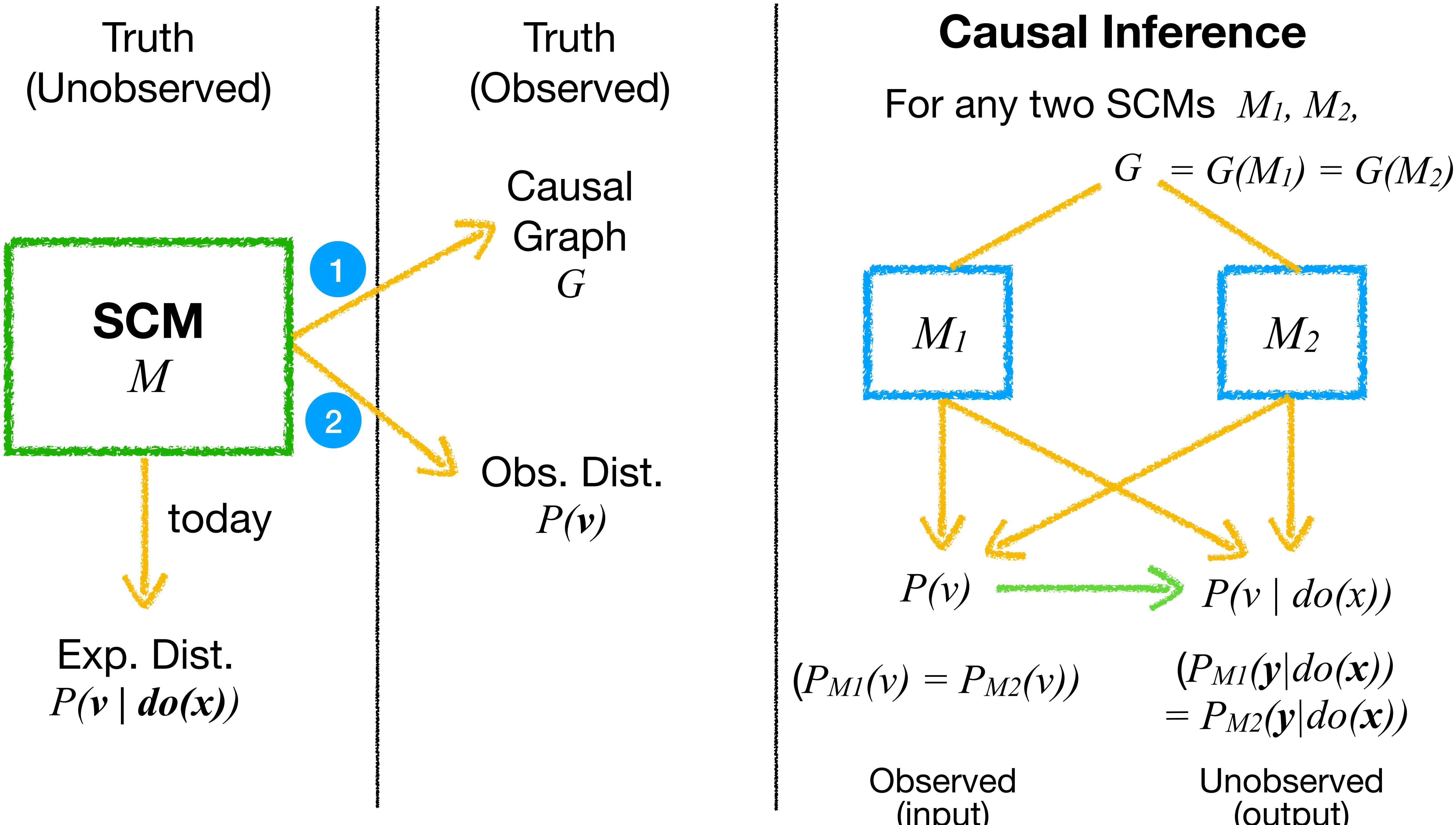
The Identification Problem

Causal Effect Identifiability (Def. 3.2.2)

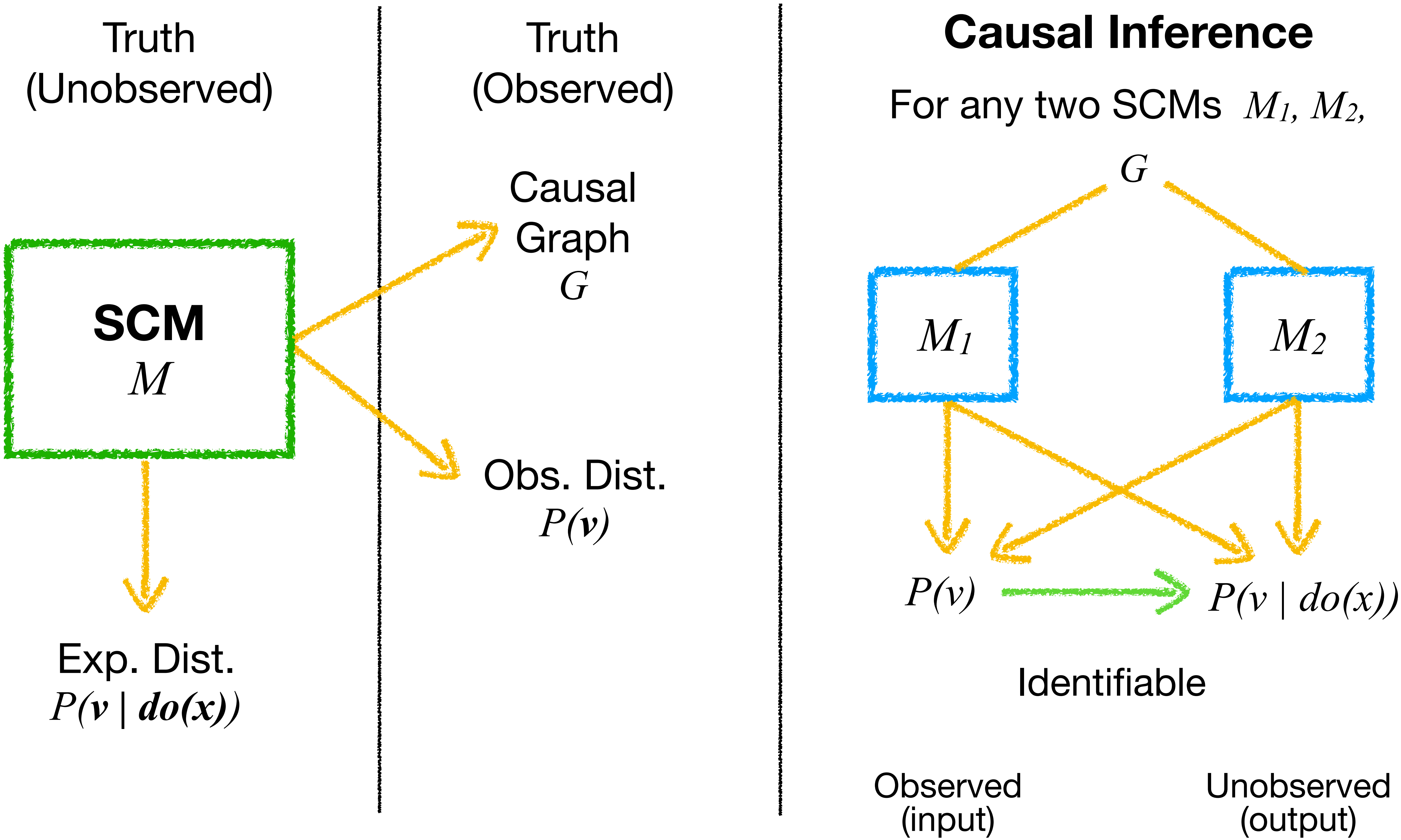
The causal effect of $\boldsymbol{X}$ on $\boldsymbol{Y}$ is said to be identifiable from a causal diagram G if the quantity $P(\mathbf{y} \mid do(\mathbf{x}))$ can be computed uniquely from a positive probability of the observed variables.

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)

SCM
M

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

Truth
(Observed)

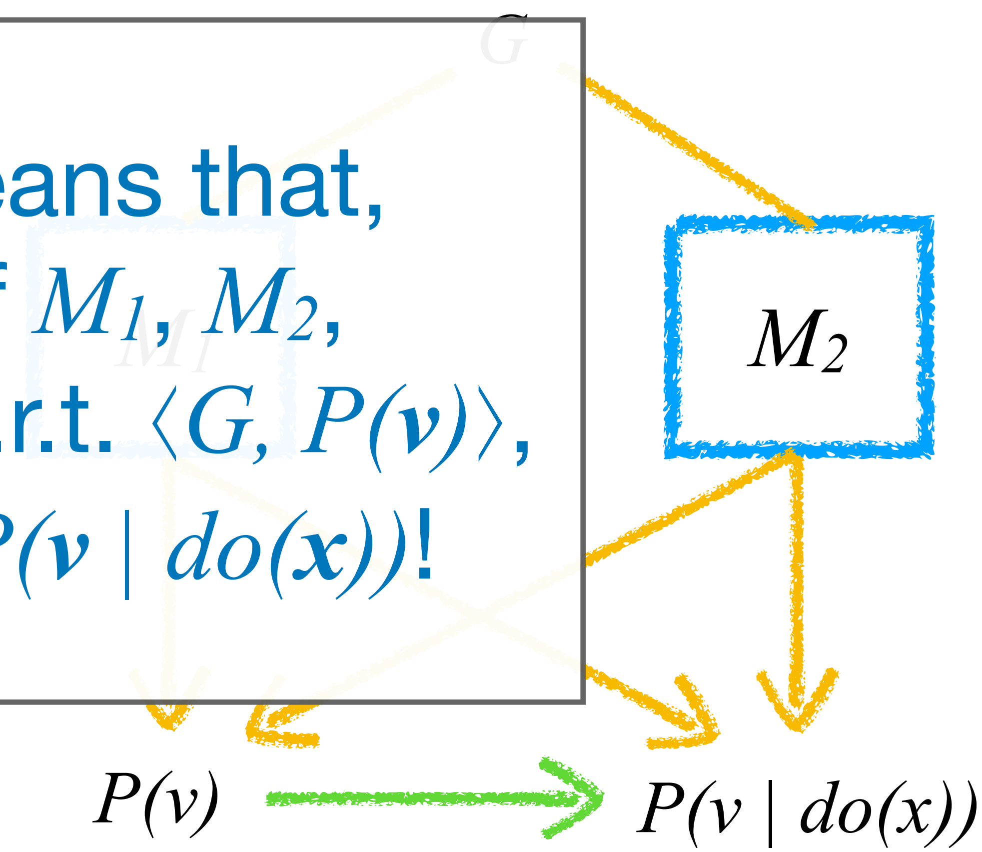
Causal Graph
G

Obs. Dist.
$P(\mathbf{v})$

Causal Inference

For any two SCMs M_1, M_2,

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

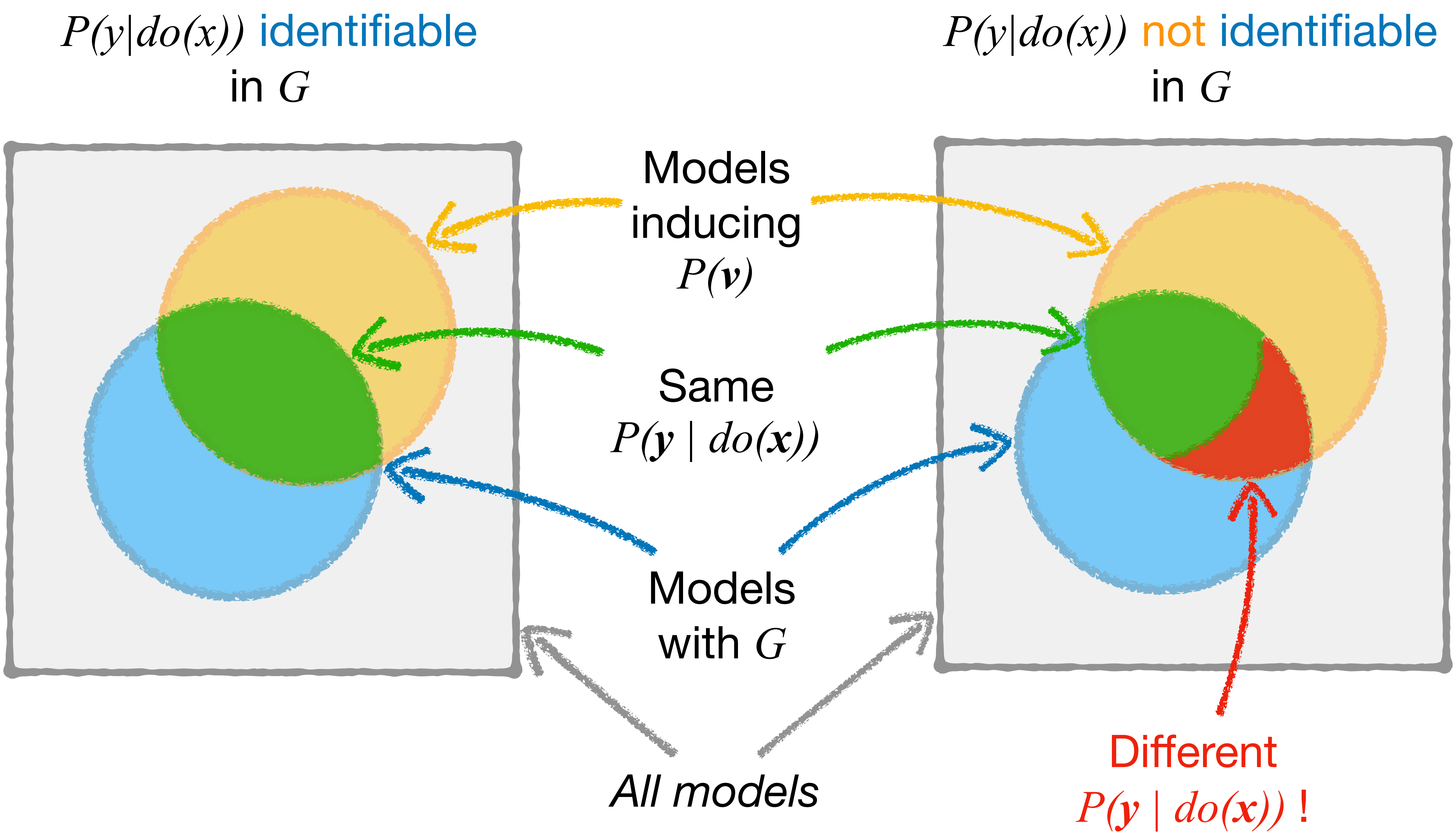


Identifiable

Observed
(input)

Unobserved
(output)

The Identification Problem (III)



Now, let's study some identifiable and non-identifiable effects…

Identification in Markovian Models

<u>Thm 4.2.1.</u> Given the causal diagram G of any Markovian model (all relevant vars are measured), the causal effect $Q = P(\boldsymbol{y} \mid do(\boldsymbol{x}))$ is identifiable for every subsets of variables $\boldsymbol{X}$ and $\boldsymbol{Y}$ and is obtained from the truncated factorization, i.e.,

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

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

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

Adjustment by Direct Parents

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

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

If Obesity is latent, is the effect still computable?

Queries:

$Q_2 = P(dm \mid do(IR = on))$

$$= \frac{\sum_{ob,htn} P'(dm \mid IR = t, htn) P'(IR = t) P'(htn \mid ob) P'(ob)}{P'(IR = t)}$$

equal to 1

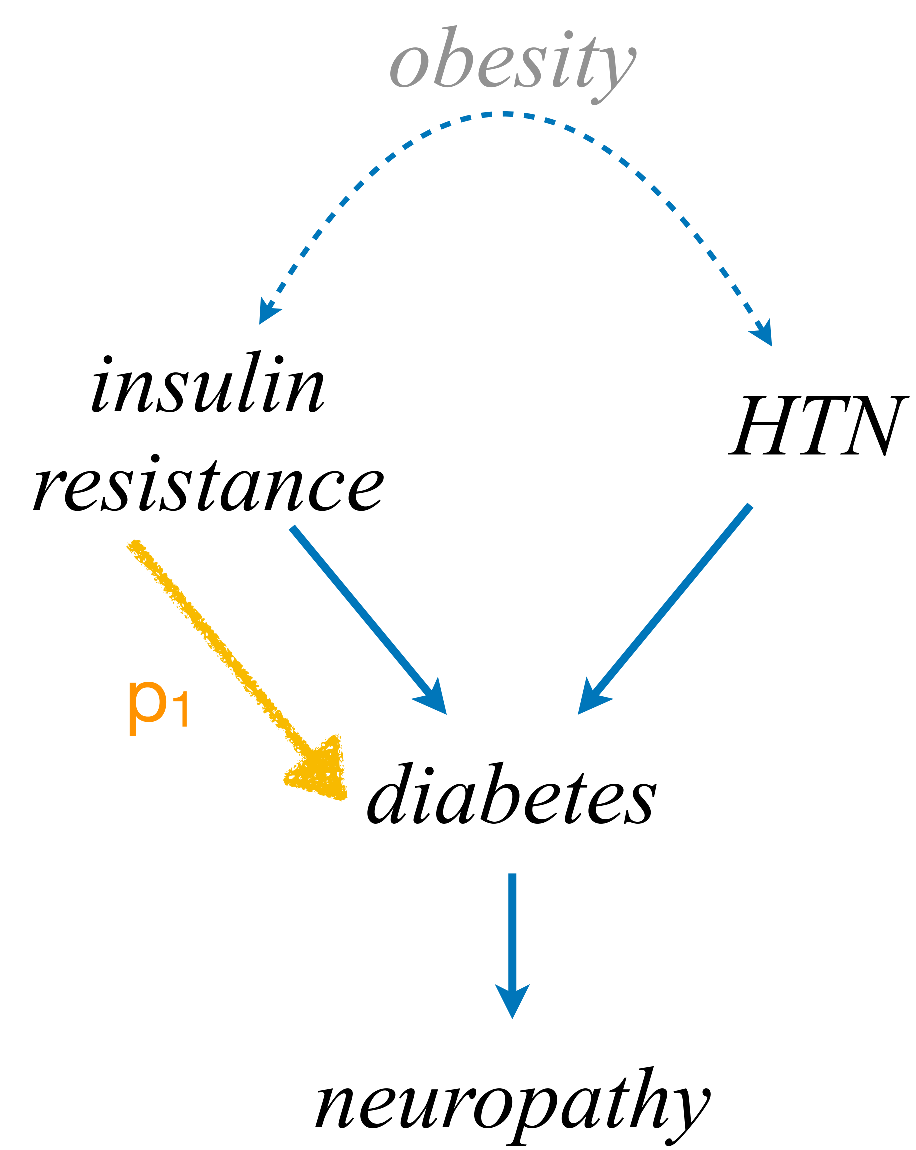
$$= \sum_{ob,htn} P(dm \mid IR = t, htn) P(htn \mid ob) P(ob)$$

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

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

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

Adjustment by hypertension!

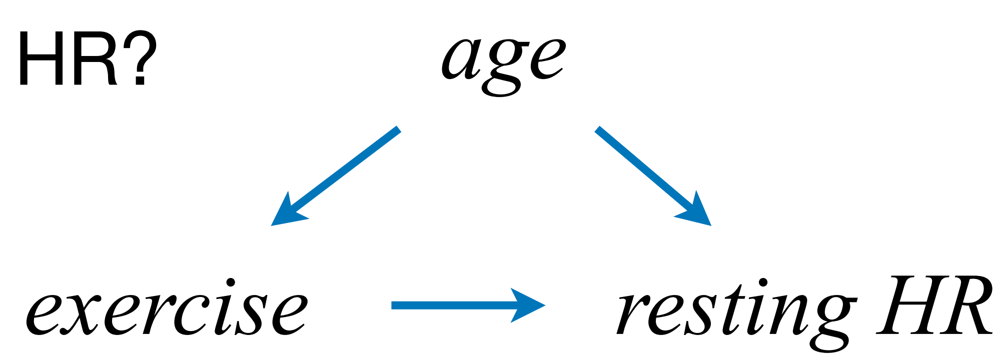


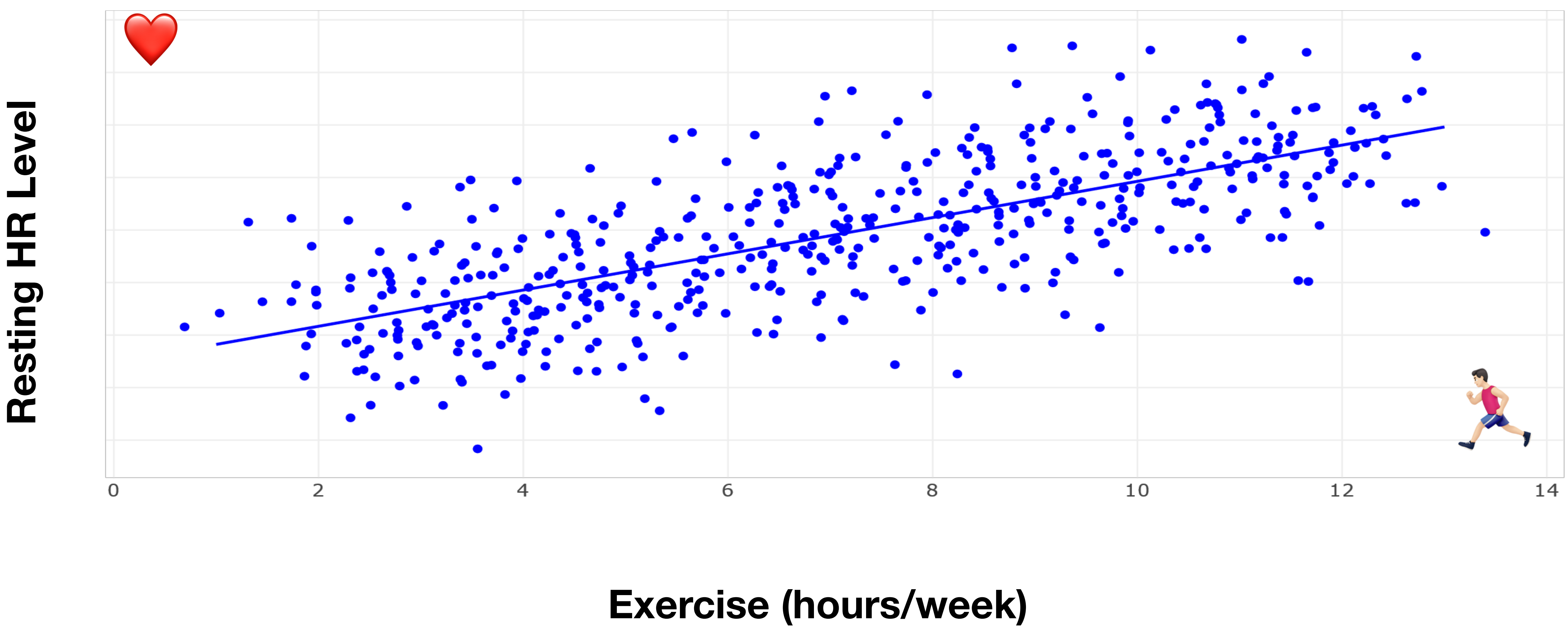
Confounding Bias

Confounding Bias

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

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





Confounding Bias

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

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