Causal Inference for Health Data

(STATS C160/C260 — Winter 2026)

Lecture 3: Testable Implications & the d-separation criterion

Drago Plečko

UCLA College | Physical Sciences
Statistics & Data Science

MIMIC-IV Data Access

SECTION ONE: DETERMINING THE TRAINING REQUIRED In order to determine which course(s) you may need to take, please indicate which of the following categories apply to you (check all that apply)

This question is required. Choose all that apply.

- [x] I conduct human subjects research/Realizo investigaciones con sujetos humanos "OHRPP" **[SECTION TWO]**
- [ ] I conduct animal research **[SECTION THREE]**
- [ ] I conduct research that involves using hazardous biological material or recombinant or synthetic nucleic acids (IBC) **[SECTION FOUR]**
- [ ] I conduct research involving human gene transfer/recombinant DNA **[SECTION FIVE]**
- [ ] I conduct research that meets the definition of Dual Use Research of Concern ("DURC") **[SECTION SIX]**
- [ ] I was instructed to take a course by Research Policy and Compliance in order to meet the Responsible Conduct of Research training requirements of the agency funding the research **[SECTION SEVEN]**
- [ ] I have a special role (OHRPP Staff, IRB member, IACUC member, IBC member, DURC committee member, Institutional Official) in the research enterprise at UCLA that requires me to take training. **[SECTION EIGHT]**
- [ ] I am looking to enroll in the **UCLA HIPAA** or **Information Privacy & Security (IPS)** courses. **[SECTION NINE]**
- [ ] I am required to take Research Security Training to meet the training requirements of the agency funding the research **[SECTION 10]**
- [ ] I am looking to enroll in a webinar. **[SECTION 11]**

Tick only the first box!

3. SCM → Pearl's Causal Hierarchy

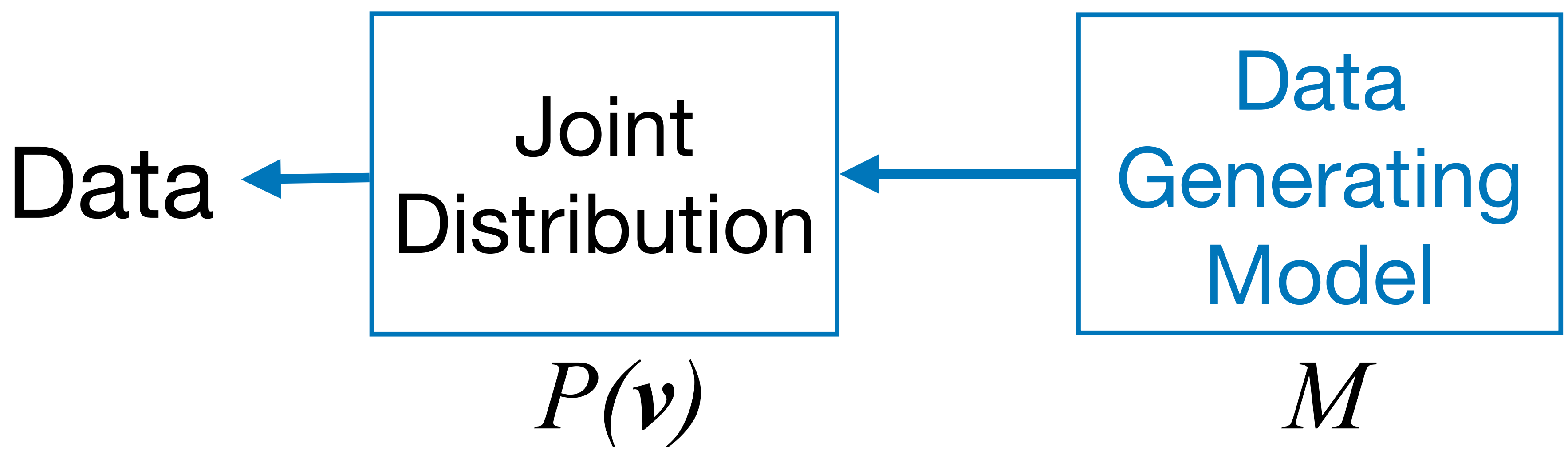
Level (Symbol)	Typical Activity	Typical Question	Examples
1 👀 Associational $P(y \mid x)$	Seeing (Associations) ML: predictive analytics	What is? How would seeing X change my belief in Y?	Is low blood pressure associated with sepsis?
2 💪 Interventional $P(y \mid do(x), c)$	Doing (Treatment Effects, Interventions) Reinforcement Learning	What if? What if I do X?	Do statins prevent heart attacks?
3 🧠 Counterfactual $P(y_x \mid x', y')$	Imagining, Retrospection (Treatment Harm)	Why? What if I had acted differently?	Had the patient not undergone surgery, would they be alive?

1st Layer of the Causal Hierarchy

— Associations

(What if I see $X=x$?)

The Emergence of the First Layer



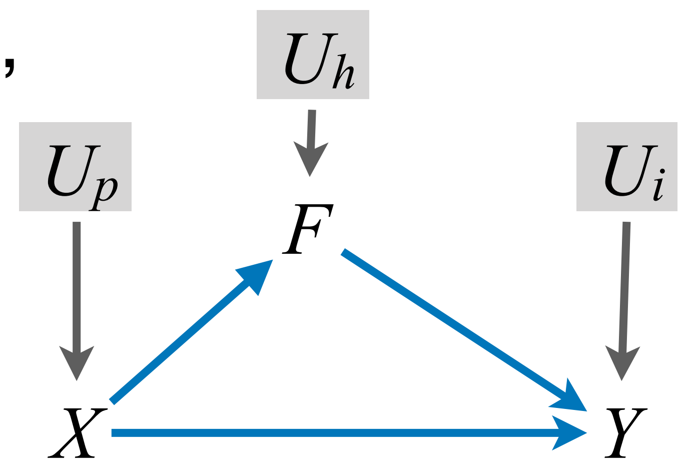
M induces a probability distribution $P(v_1, \dots, v_n)$ over the set of observable variables V. $P(v)$ can be written as follows:

$$P(\mathbf{V}) = \sum_{\mathbf{u}} P\left(\mathbf{V} = \mathbf{v}, \mathbf{U} = \mathbf{u}\right)$$
$$= \sum_{\mathbf{u}} P\left(\forall_{V_i \in \mathbf{V}}\ f_i(pa_i, u_i) = v_i, \mathbf{U} = \mathbf{u}\right)$$

where pa_i and u_i are the (values of) the observed and unobserved arguments of f_i, which are the parents of V_i in the corresponding causal diagram G.

The Emergence of the First Layer

In the previous example,



The joint distribution over the observables $P(v)$ is equal to:

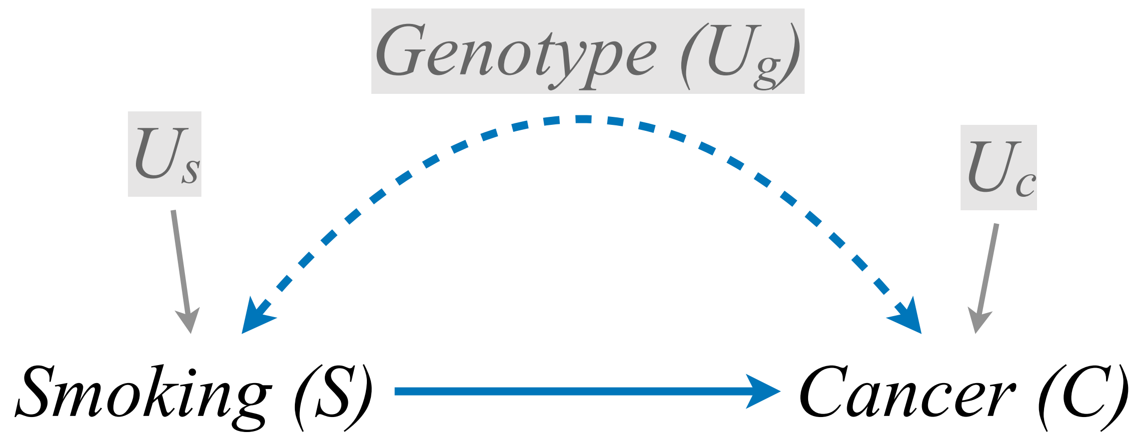
$$P(X = x, F = f, Y = y) = \sum_{u_p, u_h, u_i} P(X = x, F = f, Y = y, U_p = u_p, U_h = u_h, U_i = u_i)$$

For short,

$$P(x, f, y) = \sum_{u_p, u_h, u_i} P(x, f, y, u_p, u_h, u_i)$$

The Emergence of the First Layer

In the second example,



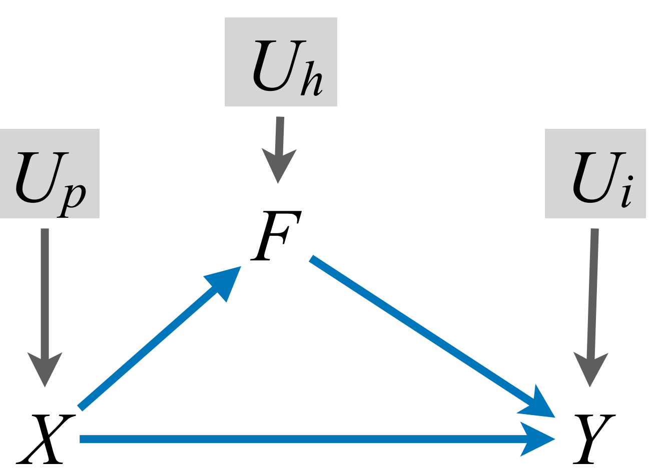
The joint probability distribution over the observed variables (V), *Smoking* and *Cancer*, is given by

$$P(s, c) = \sum_{u_s, u_g, u_c} P\left(s, c, u_s, u_g, u_c\right)$$

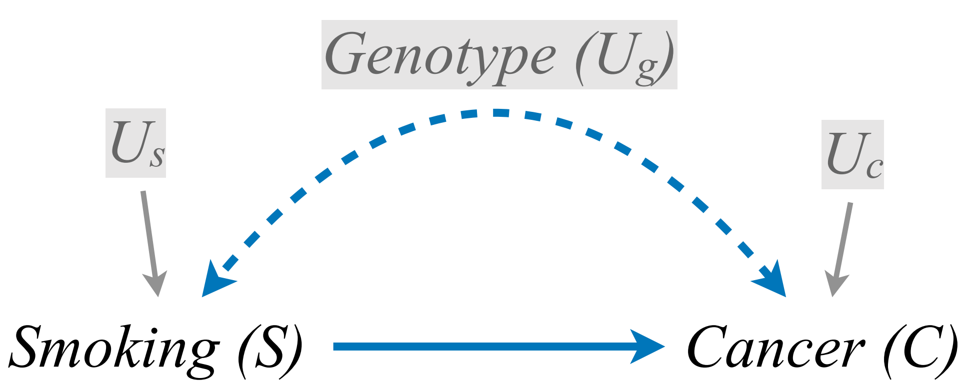
Recall, the l.h.s. distribution is called observational distribution. Sometimes, it's also called passive or non-experimental distribution.

What the diagram encodes

- Since G is a directed acyclic graph, there exists a topological order over V such that every variable goes after its parents, i.e., $Pa_i < V_i$.



$$X < F < Y$$



$$S < C$$

What the diagram encodes

- SCM M induces $P(V)$:

$$P(\mathbf{v}) = \sum_{\mathbf{u}} P(\mathbf{v}, \mathbf{u}),$$

- Using the chain rule and following a topological order,

$$P(\mathbf{v}) = \sum_{\mathbf{u}} P(\mathbf{u}) \prod_{V_i \in \mathbf{V}} P\left(v_i \mid v_1, \ldots, v_{i-1}, \mathbf{u}\right),$$

- An observed variable is fully determined by its observed and unobserved parents;

$$P\left(v_i \,|\, v_1, \ldots, v_{i-1}, \mathbf{u}\right) = P\left(f_{V_i}(pa_i, u_i) = v_i \,|\, v_1, \ldots, v_{i-1}, \mathbf{u}\right)$$

also $\{pa_i, u_i\} \subseteq \{v_1, \ldots, v_{i-1}, \boldsymbol{u}\}$, then

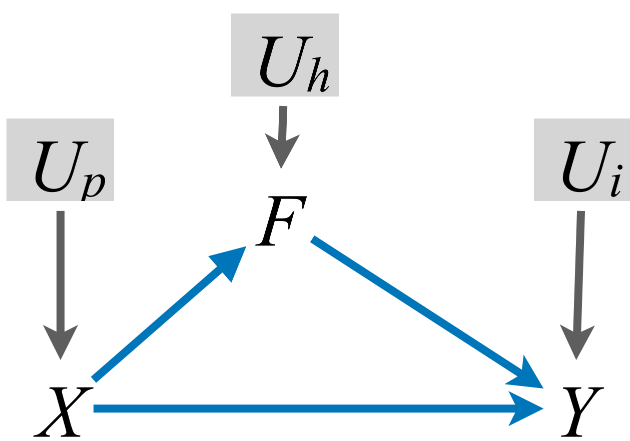
$$\boxed{P\left(v_i \,|\, v_1, \ldots, v_{i-1}, \mathbf{u}\right) = P\left(v_i \,|\, pa_i, u_i\right)}$$

What the diagram encodes

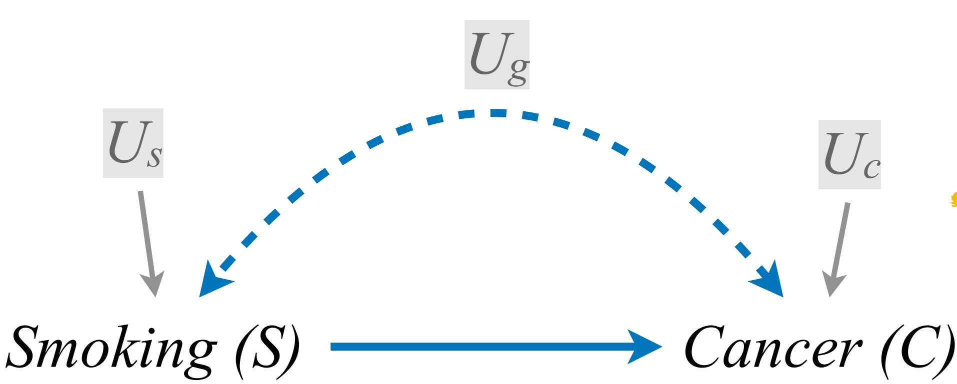
- The distribution *P(V)* decomposes as:

$$P(\mathbf{v}) = \sum_{\mathbf{u}} P(\mathbf{u}) \prod_{V_i \in \mathbf{V}} P\left(v_i \mid v_1, \ldots, v_{i-1}, \mathbf{u}\right)$$

$$= \sum_{\mathbf{u}} P(\mathbf{u}) \prod_{V_i \in \mathbf{V}} P\left(v_i \mid pa_i, u_i\right)$$



$$P(r, d, a) = \sum_{u_p, u_h, u_i} P(u_p, u_h, u_i) P(x \mid u_p) P(f \mid x, u_h) P(y \mid x, f, u_i)$$



$$P(s, c) = \sum_{u_s, u_g, u_c} P(u_s, u_g, u_c) P(s \mid u_g, u_s) P(c \mid s, u_g, u_c)$$

Probabilistic Invariance
Conditional Independences

- If knowing that variable X = x doesn't change the belief in Y = y, then X and Y are said to be probabilistically independent. This is written as $X \perp\!\!\!\perp Y$.
  - $X \perp\!\!\!\perp Y \equiv P(Y = y \mid X = x) = P(Y = y)$

$P(Y = y, X = x) = P(Y = y)\, P(X = x)$

- More generally, once we know the value of a third variable Z = z, if knowing that X = x doesn't affect the belief of Y = y, X and Y are conditionally independent given Z, i.e., $X \perp\!\!\!\perp Y \mid Z$.
  - $X \perp\!\!\!\perp Y \mid Z \equiv P(Y = y \mid Z = z, X = x) = P(Y = y \mid Z = z).$
- These are properties of (constraints over) *P(V).*

Lack of functional dependence → probabilistic independence.

G → Cond. Independences in P

- Recall that in any topological order:

$$P\left(v_i \,|\, v_1, \ldots, v_{i-1}, \mathbf{u}\right) = P\left(v_i \,|\, pa_i, u_i\right)$$

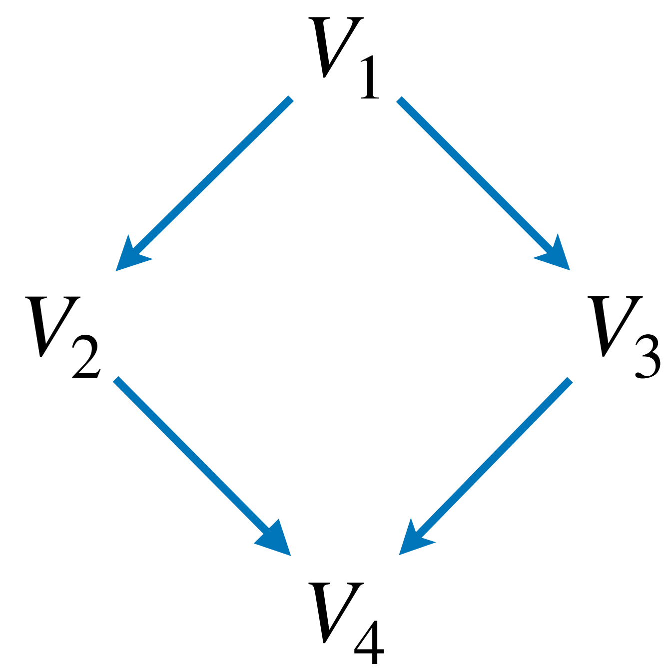
- What is the maximal set $\{V_1, \ldots, V_{i-1}\}$ in any topological order?
    - It may include any variable W as long as there is no chain

$$V_i \rightarrow \cdots \rightarrow \cdots \rightarrow W$$

This includes all the <u>non-descendants</u> of V_i, denote as Nd_i.

Example: Non-Descendants

- $Nd(V_1) = \emptyset$,
- $Nd(V_2) = \{V_1, V_3\}$,
- $Nd(V_3) = \{V_1, V_2\}$,
- $Nd(V_4) = \{V_1, V_2, V_3\}$



G → Cond. Independences in P

- Then, we have:

$$P\left(v_i \,|\, v_1, \ldots, v_{i-1}, \mathbf{u}\right) = P\left(v_i \,|\, pa_i, u_i\right)$$

$$P\left(v_i \,|\, nd_i, \mathbf{u}\right) \overset{(*)}{=} P\left(v_i \,|\, pa_i, u_i\right)$$

- This is expressing an independence relation,

rewrite LHS of (*)

$$\frac{P\left(v_i, nd_i \backslash pa_i, \mathbf{u} \backslash u_i \,|\, pa_i, u_i\right)}{P\left(nd_i \backslash pa_i, \mathbf{u} \backslash u_i \,|\, pa_i, u_i\right)} = P\left(v_i \,|\, pa_i, u_i\right)$$

$$P\left(v_i, nd_i \backslash pa_i, \mathbf{u} \backslash u_i \,|\, pa_i, u_i\right) = P\left(v_i \,|\, pa_i, u_i\right) P\left(nd_i \backslash pa_i, \mathbf{u} \backslash u_i \,|\, pa_i, u_i\right)$$

- In other words, $(V_i \perp\!\!\!\perp Nd_i \setminus Pa_i, \boldsymbol{U} \setminus U_i \mid Pa_i, U_i)$

- This, in turn, implies $(V_i \perp\!\!\!\perp Nd_i \setminus Pa_i \mid Pa_i, U_i)$

Markovian Factorization

- If no variable in U is a parent of more than one variable in V *(observables)* (i.e., $\forall_{i,j}\ U_i \cap U_j = \emptyset$), then the model is called Markovian. We have:

$$
\begin{aligned}
P(\mathbf{v}) &= \sum_{\mathbf{u}} P(\mathbf{u}) \prod_{V_i \in \mathbf{V}} P\left(v_i \mid pa_i, u_i\right) \\
&= \sum_{\mathbf{u}} \prod_{V_i \in \mathbf{V}} P\left(v_i \mid pa_i, u_i\right) P\left(u_i\right) \\
&= \sum_{\mathbf{u}} \prod_{V_i \in \mathbf{V}} P\left(v_i \mid pa_i, u_i\right) P\left(u_i \mid pa_i\right) \\
&= \prod_{V_i \in \mathbf{V}} \sum_{u_i} P\left(v_i, u_i \mid pa_i\right) = \boxed{\prod_{V_i \in \mathbf{V}} P\left(v_i \mid pa_i\right)}
\end{aligned}
$$

Local Markovian Condition

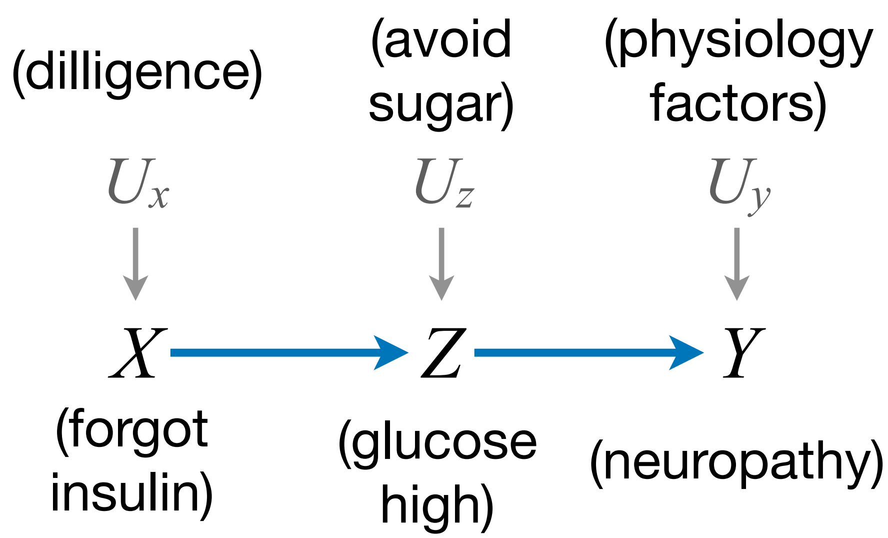
$$\boxed{(V_i \perp\!\!\!\perp Nd_i \setminus Pa_i \mid Pa_i)}$$

Bayesian Factorization

Connecting Unobserved and Observed Worlds

(or, Implications of the Markov Condition)

Causal Chains



Are X and Y independent?

No,

Knowing $X{=}1$ (insulin not taken) changes the neuropathy likelihood ($Y{=}1$).

$$P(X=1, Y=1) = \sum_z P(X=1)\,P\left(z \mid X=1\right)\,P\left(Y=1 \mid z\right)$$
$$= P(X=1)\sum_z P\left(z \mid X=1\right)\,P\left(Y=1 \mid z\right)$$
$$= P(U_x=1)P(U_y=1)$$

$$P(X=1)P(Y=1)$$
$$= P(X=1)\sum_z P(Y=1 \mid z)P(z)$$
$$= P(U_x=1)\sum_z P(U_y \wedge z=1 \mid z)P(z)$$
$$= P(U_x=1)P(U_y=1)P(Z=1)$$
$$= P(U_x=1)P(U_y=1)\sum_x P(Z=$$

Pick $P(U_z{=}0) \neq 1$.

$$= P(U_x=1)P(U_y=1)\left(P(U_z=0)P(U_x=0)+P(U_x=1)\right)$$

M: $X \leftarrow U_x$

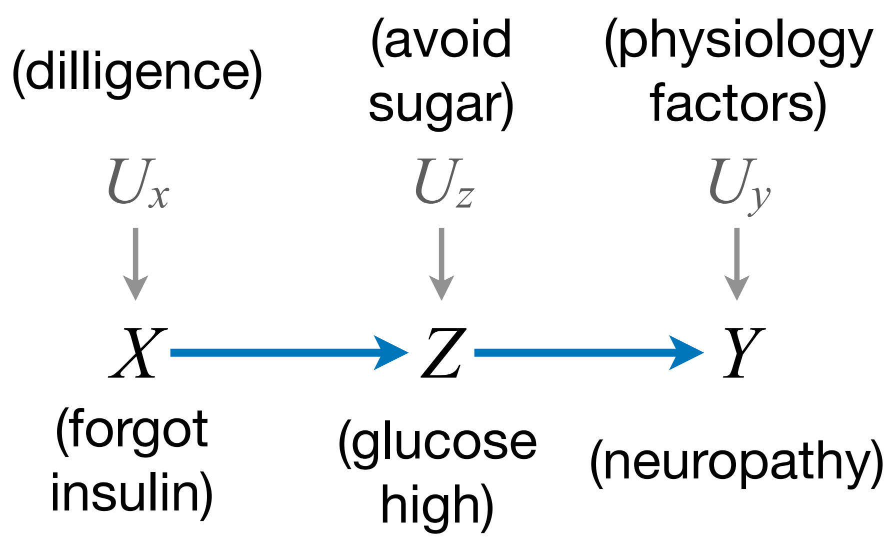
$Z \leftarrow X \vee \neg U_z$

$Y \leftarrow Z \wedge U_y$

$P(U_x, U_z, U_y)$

Graphically, information "flows" from X to Y through Z.

Causal Chains



Are X and Y independent given Z?

Yes,

e.g., knowing $Z=1$ (high glucose), the probability of neuropathy ($Y=1$) does not change if we know insulin wasn't taken ($X=1$) or was ($X=0$).

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

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

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

Bayes Factorization

$$M: \; X \leftarrow U_x$$

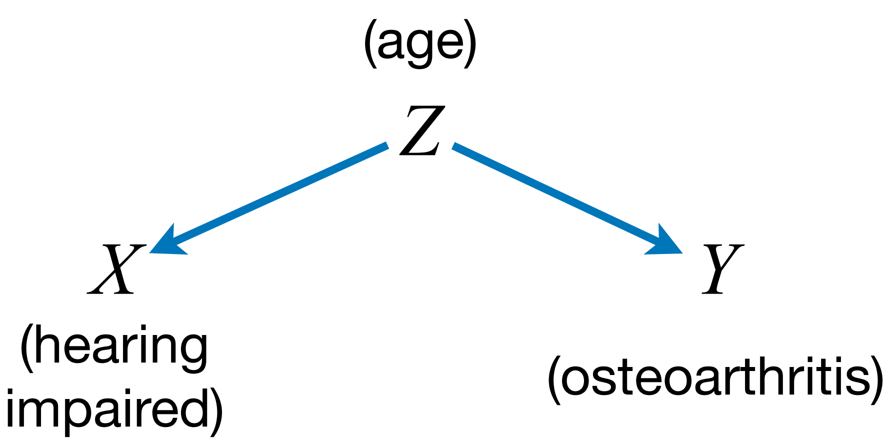
$$Z \leftarrow X \vee \neg U_z$$

$$Y \leftarrow Z \wedge U_y$$

$$\mathrm{P}(U_x, U_z, U_y)$$

Graphically, observing Z "blocks" the influence from X to Y.

Common Cause



Are X and Y independent?

No,

e.g., seeing someone with osteoarthritis ($Y=1$) raises the probability of older age ($Z=1$), which increases the likelihood of impaired hearing ($X=1$).

$$\exists_{x,y} P(X=x, Y=y) \neq P(X=x)P(Y=y)$$

try it out!

Graphically, information "flows" from Y going through the common cause Z and down to X.

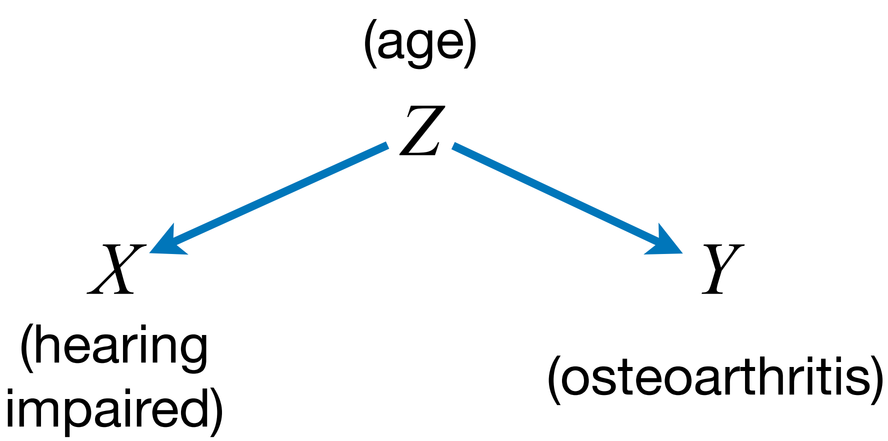
M: $Z \leftarrow U_z$

$X \leftarrow Z \vee U_x$

$Y \leftarrow Z \vee U_y$

$P(U_x, U_z, U_y)$

Common Cause



Are X and Y independent given Z?

Yes,

e.g., if we know the patient is old ($Z=1$), observing osteoarthritis ($Y=1$) tell us nothing about the impaired hearing (X).

Bayes Factorization

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

Graphically, observing Z "blocks" the influence from X to Y.

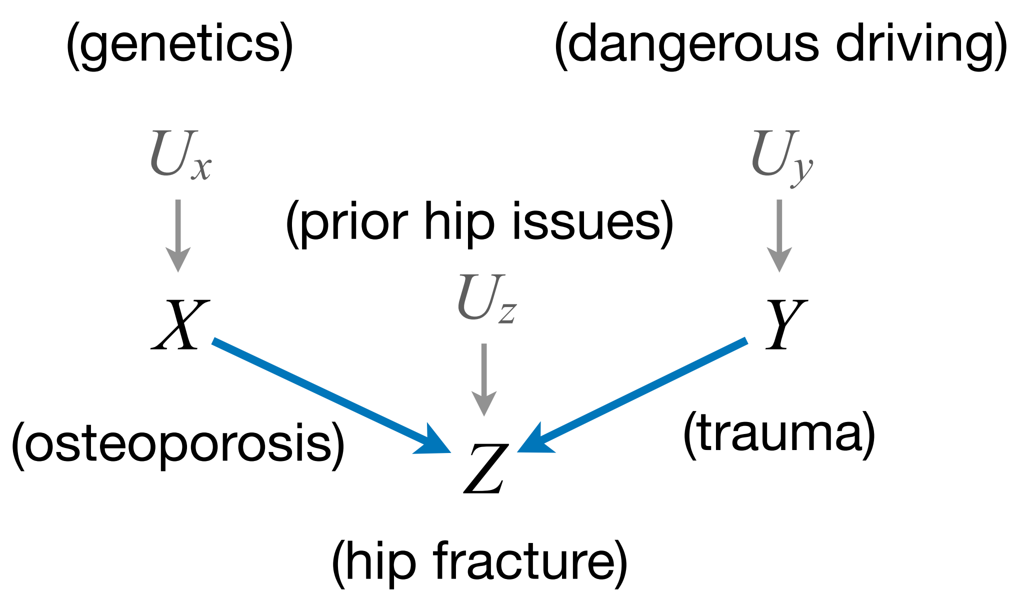
M: $Z \leftarrow U_z$

$X \leftarrow Z \vee U_x$

$Y \leftarrow Z \vee U_y$

$P(U_x, U_z, U_y)$

Common Effect



Are X and Y independent?

Yes!,

e.g., having osteoporosis $X=1$ is independent of trauma $Y=1$.

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

Graphically, influence from X reaches Z but does not "go up" to Y.

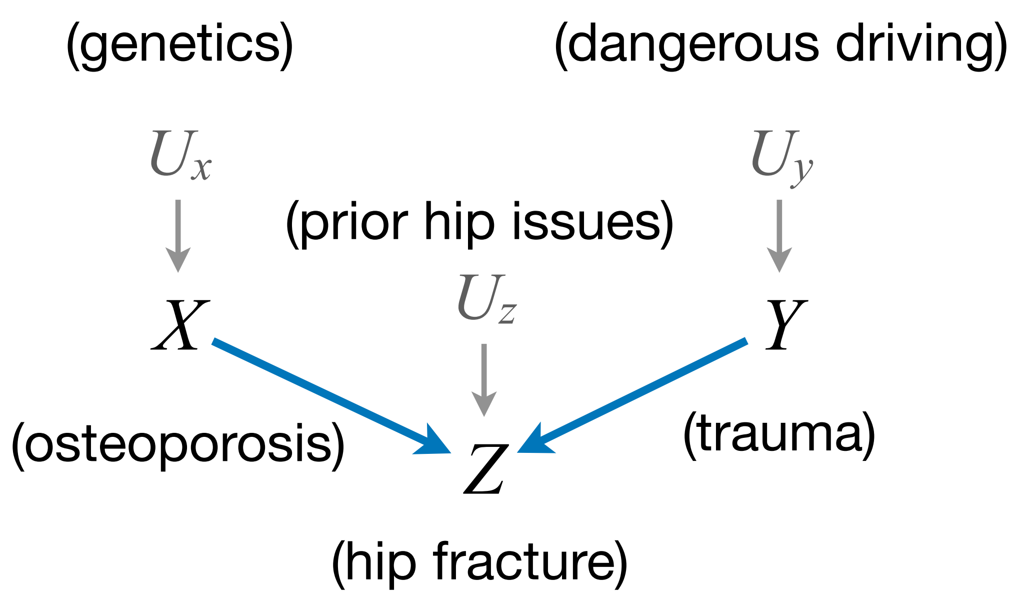
M: $X \leftarrow U_x$

$Y \leftarrow U_y$

$Z \leftarrow Y \wedge (X \vee U_z)$

$P(U_x, U_z, U_y)$

Common Effect



Are X and Y independent given Z?

No!

e.g., if we observe that a patient has a hip fracture ($Z=1$) but did not have trauma ($Y=0$), it is more likely that they have osteoporosis ($X=1$). Somethings needs to "explain" the observed fracture.

$$\exists_{x,y,z} P(X = x, Y = y \mid Z = z) \neq P(X = x \mid Z = z) P(Y = y \mid Z = z)$$

try it out!

M: $X \leftarrow U_x$

$Y \leftarrow U_y$

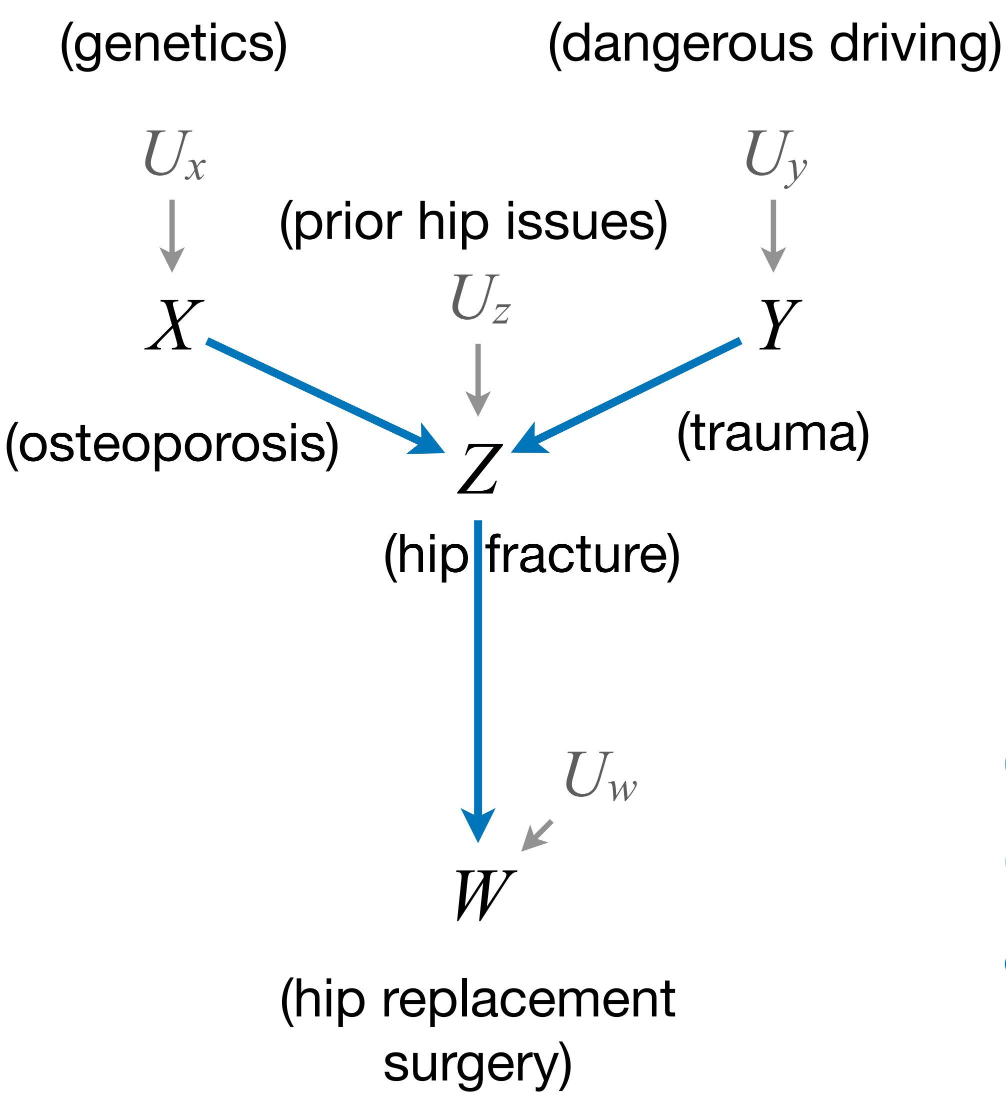
$Z \leftarrow Y \wedge (X \vee U_z)$

$P(U_x, U_z, U_y)$

Graphically, influence from X reaching Z (when Z is observed) bumps "back up" to Y.

This behavior is the opposite of the previous cases!

Common Effect



Are X and Y independent given W?

No!, again!,

e.g., observing that a patient had a hip replacement surgery ($W=1$) increases the likelihood that they had a fracture ($Z=1$), and from before, this makes X and Y dependent.

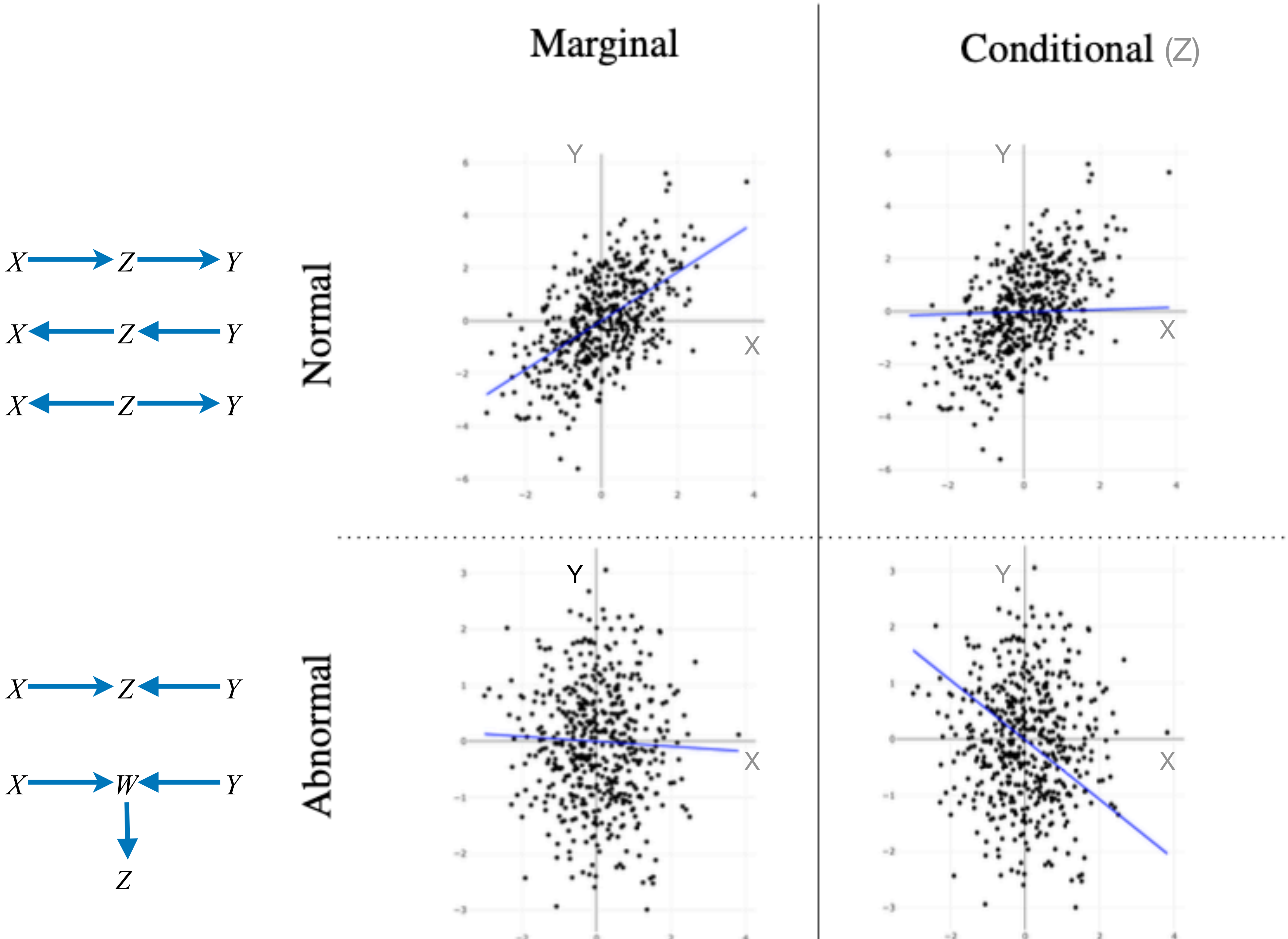
Graphically, influence from X reaching W (when W is observed) "bumps back up" to Z, and then Y.

Watch out for the descendants of the colliders!

Summary

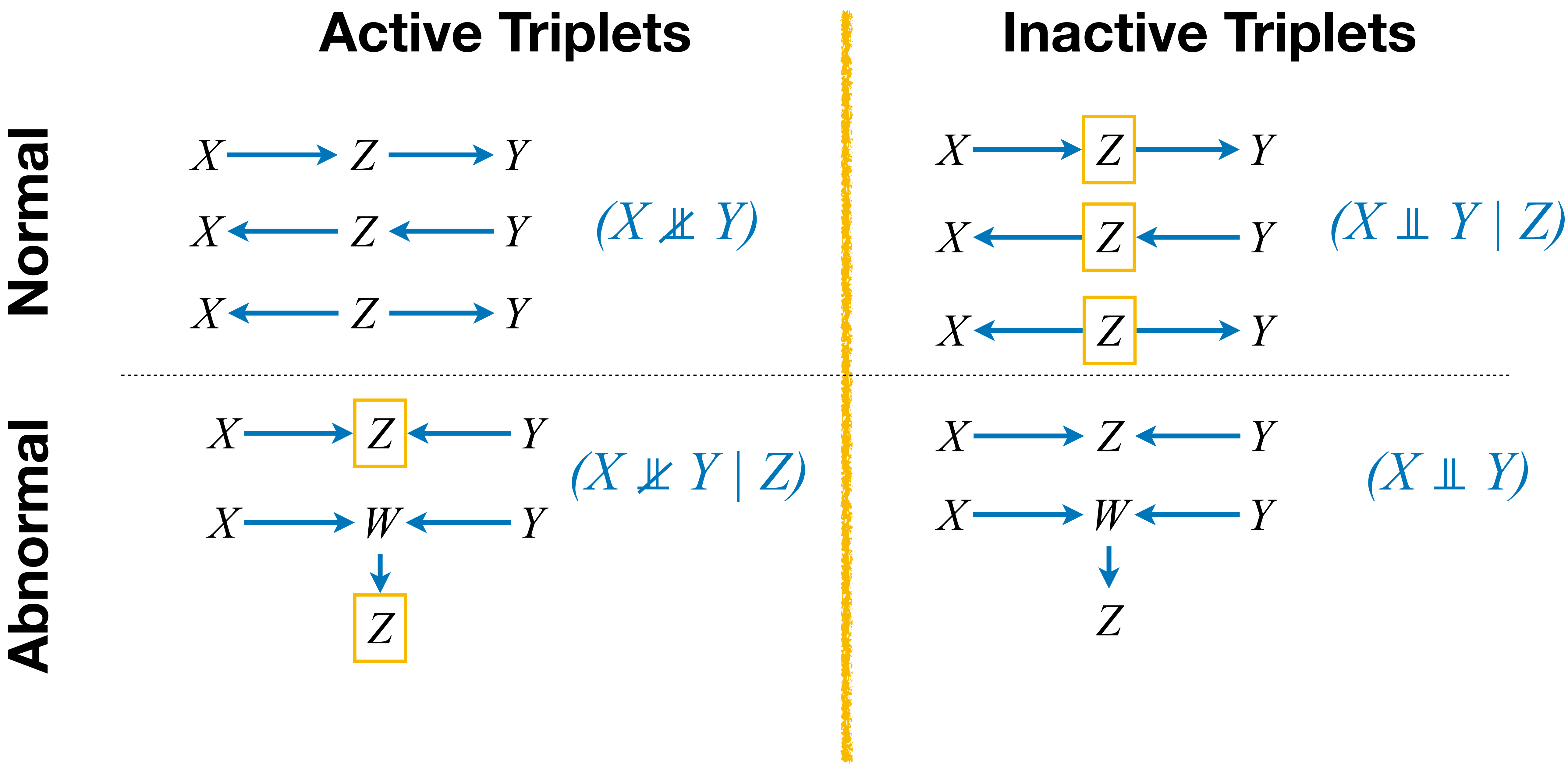
		Marginal	Conditional (Z)
$X \rightarrow Z \rightarrow Y$ $X \leftarrow Z \leftarrow Y$ $X \leftarrow Z \rightarrow Y$	Normal		
	Abnormal		

Summary



Marginal
Conditional (Z)
Normal
Abnormal
X → Z → Y
X ← Z ← Y
X ← Z → Y
X → Z ← Y
X → W ← Y
Z
Y
X

Summary



What about larger graphical structures?

Graph Separation (d-Separation)

- Consider the question of whether X and Y are independent given Z.

1. Look at every path from X to Y in the graph.
2. A path is active if every triplet in it is active (given Z).
3. If any path is active,
   X and Y are not independent.

Graph Separation (d-Separation)

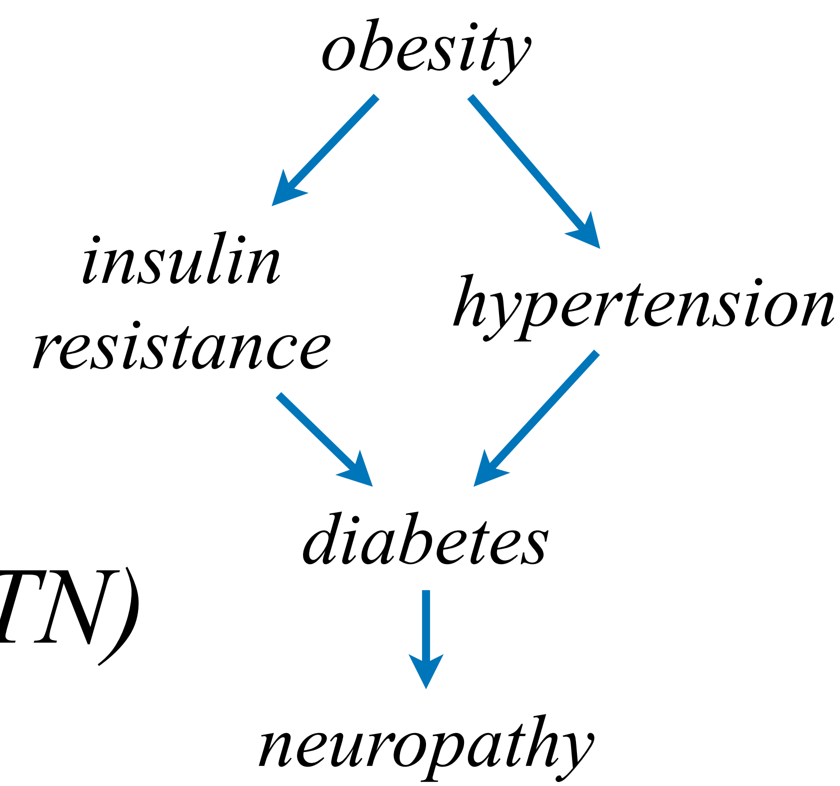
CI$_1$: $(\textit{diab} \perp\!\!\!\perp \textit{ins. resist.})$

CI$_2$: $(\textit{diab} \perp\!\!\!\perp \textit{obesity} \mid \textit{ins. resist.})$

CI$_3$: $(\textit{HTN} \perp\!\!\!\perp \textit{neuropathy} \mid \textit{diabetes})$

CI$_4$: $(\textit{obesity} \perp\!\!\!\perp \textit{diabetes} \mid \textit{ins. resist., HTN})$

CI$_5$: $(\textit{ins. resist.} \perp\!\!\!\perp \textit{HTN} \mid \textit{obesity, diab})$



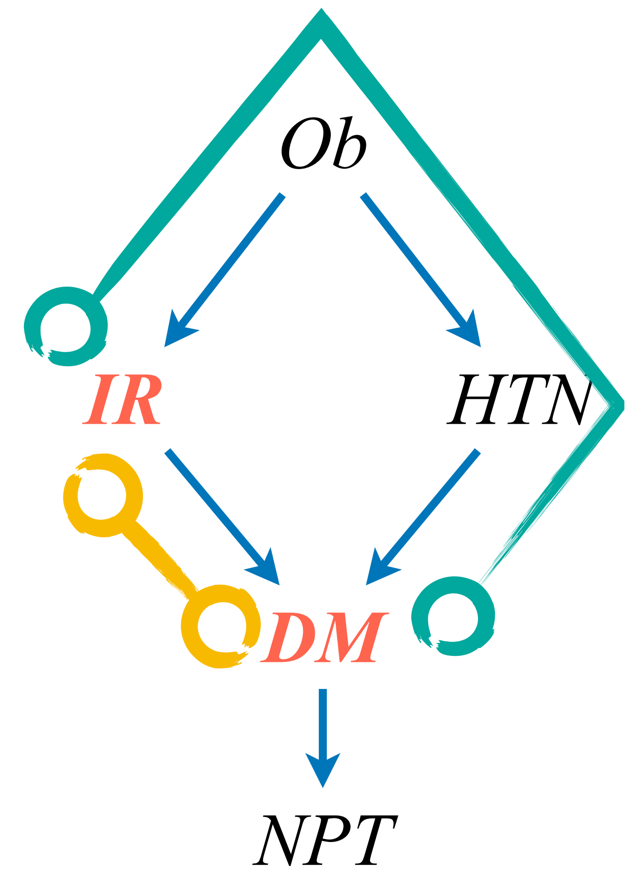
Graph Separation (d-Separation)

CI$_1$: $(\textit{diabetes} \perp\!\!\!\perp \textit{ins. resist.})$?

Path 1: $IR \longrightarrow DM$ always active

Path 2: $IR \longleftarrow Ob \longrightarrow HTN \longrightarrow We$

There exists a path that is active (actually two), hence *Insulin Resistance* and *Diabetes* are not d-separated.

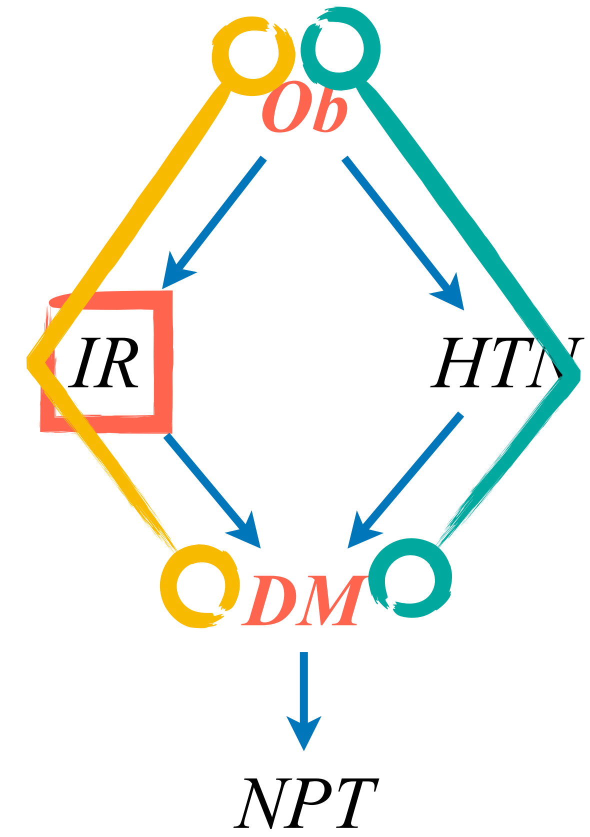


Graph Separation (d-Separation)

CI_2: $(DM \perp\!\!\!\perp Obesity \mid IR)$?

Path 1: $Ob \longrightarrow IR \longrightarrow DM$

Path 2: $Ob \longrightarrow HTN \longrightarrow DM$



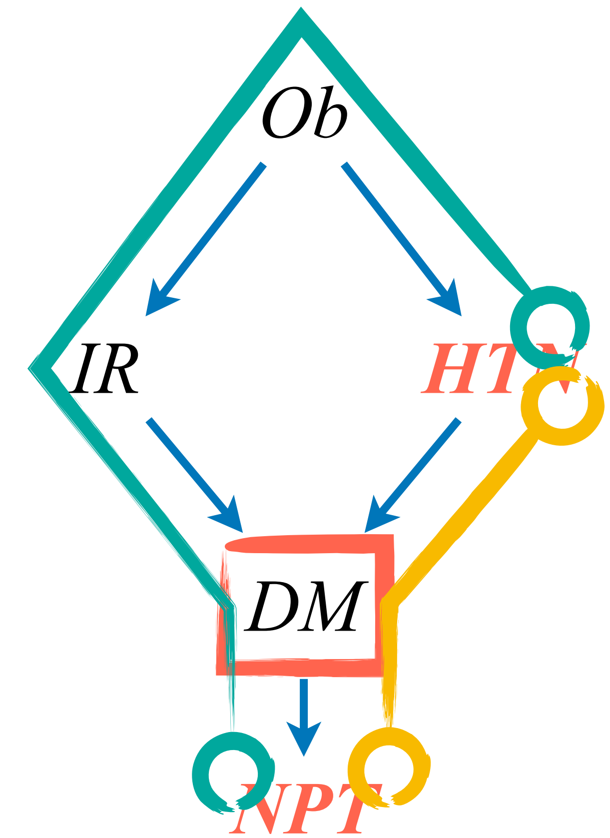
There exists a path that is active, hence *Obesity* and *Diabetes* are not d-separated given *Insulin resistance.*

Graph Separation (d-Separation)

CI3: $(HTN \perp\!\!\!\perp NPT \mid DM)$?

Path 1: $HTN \longrightarrow DM \longrightarrow NPT$

Path 2: $HTN \longleftarrow Ob \longrightarrow IR \longrightarrow DM \longrightarrow NPT$



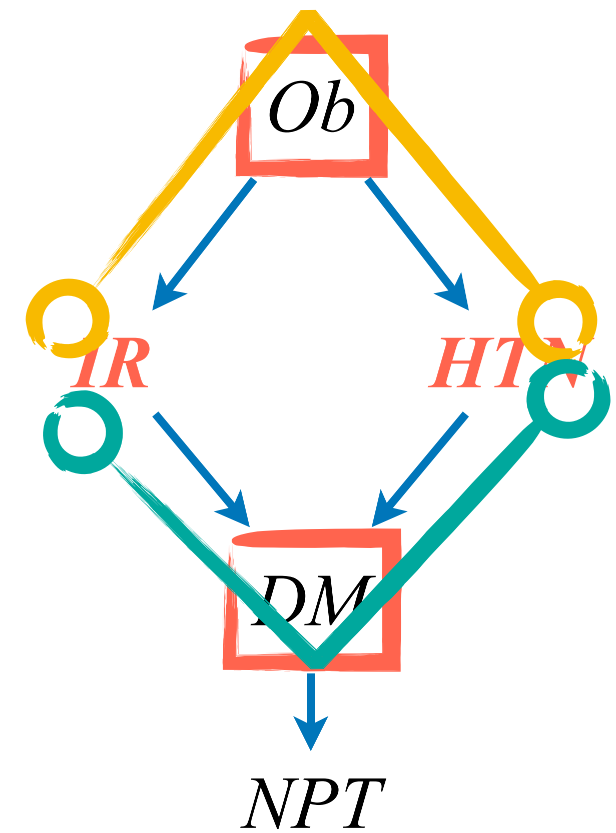
There exists **no** path that is active between *Hypertension* and *Neuropathy* given *Diabetes*, hence they are d-separated.

Graph Separation (d-Separation)

CI5: $(IR \perp\!\!\!\perp HTN \mid Ob, DM)$?

Path 1: $IR \leftarrow Ob \rightarrow HTN$

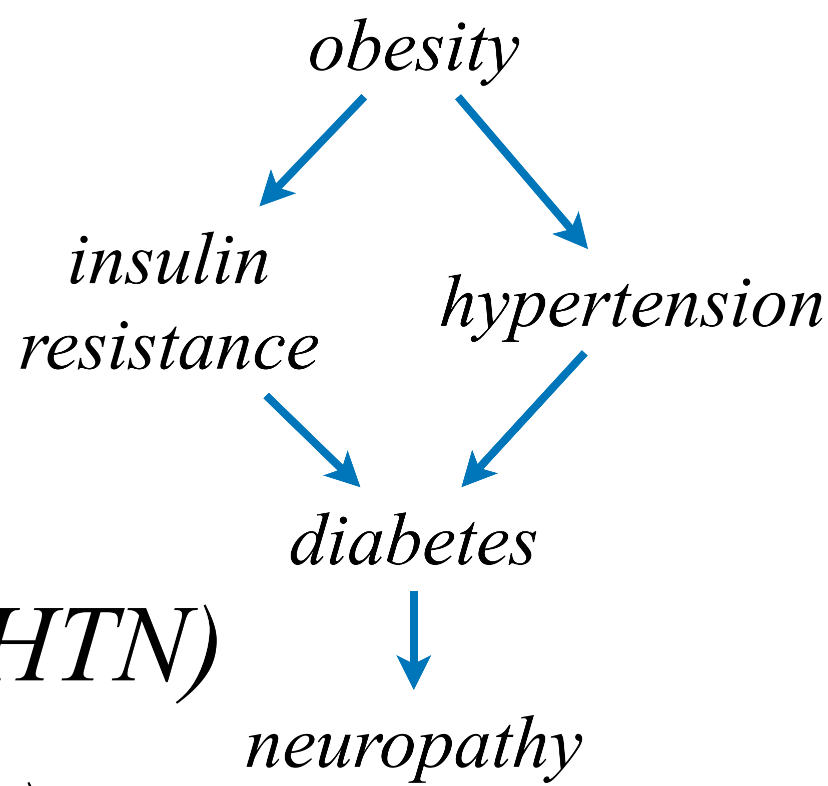
Path 2: $IR \rightarrow DM \leftarrow HTN$ becomes active given *DM*



There exists a path that is active between *Insulin Resistance* and *Hypertension* given *Obesity* and *Diabetes*, hence they are **not** d-separated.

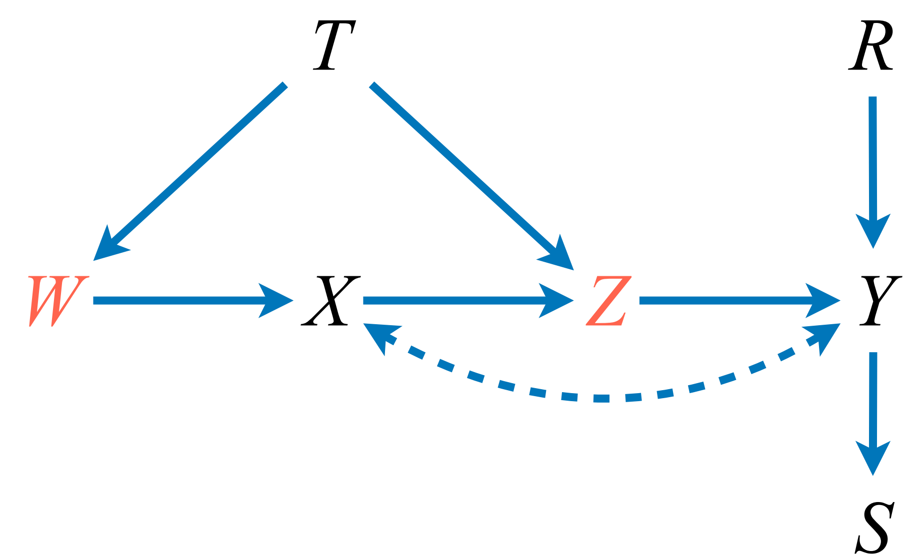
Graph Separation (d-Separation)

- ✘ CI$_1$: *(diab ⫫ ins. resist.)*
- ✘ CI$_2$: *(diab ⫫ obesity | ins. resist.)*
- ✔ CI$_3$: *(HTN ⫫ neuropathy | diabetes)*
- ✔ CI$_4$: *(obesity ⫫ diabetes | ins. resist., HTN)*
- ✘ CI$_5$: *(ins. resist. ⫫ HTN | obesity, diab)*



Graph Separation (d-Separation)

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



Path 1: $W \longleftarrow T \longrightarrow Z$

Path 2: $W \longrightarrow X \longrightarrow Z$

Path 3: $W \longrightarrow X \longleftrightarrow Y \longleftarrow Z$

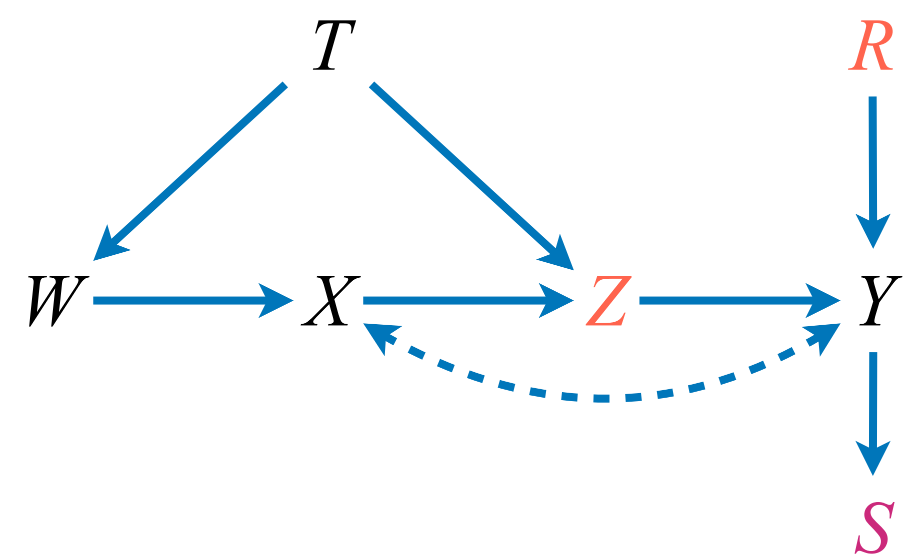
$= W \longrightarrow X \longleftarrow U \longrightarrow Y \longleftarrow Z$

Path 1 and 2 need to be blocked, Path 3 is naturally blocked:

$A = \{T, X\}$ suffices.

Graph Separation (d-Separation)

Is there a set A such that the separation statement $(R, Z \perp\!\!\!\perp S \mid A)$ holds?



Path 1: $R \longrightarrow Y \longrightarrow S$

Path 2: $Z \longrightarrow Y \longrightarrow S$

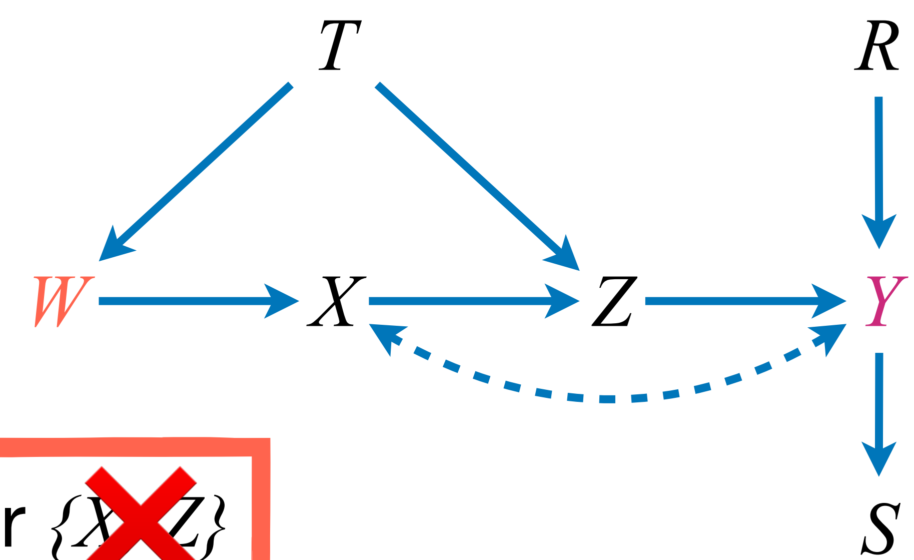
Path 3: $Z \longleftarrow X \longleftrightarrow Y \longrightarrow S$

Path 4: $Z \longleftarrow T \longrightarrow W \longrightarrow X \longleftrightarrow Y \longrightarrow S$

$A = \{Y\}$ suffices.

Graph Separation (d-Separation)

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

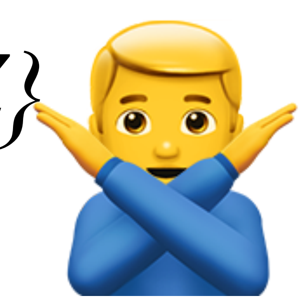


Path 1: $W \longrightarrow X \longrightarrow Z \longrightarrow Y$ ~~$\{X\}$~~ or $\{Z\}$ or ~~$\{X, Z\}$~~

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

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

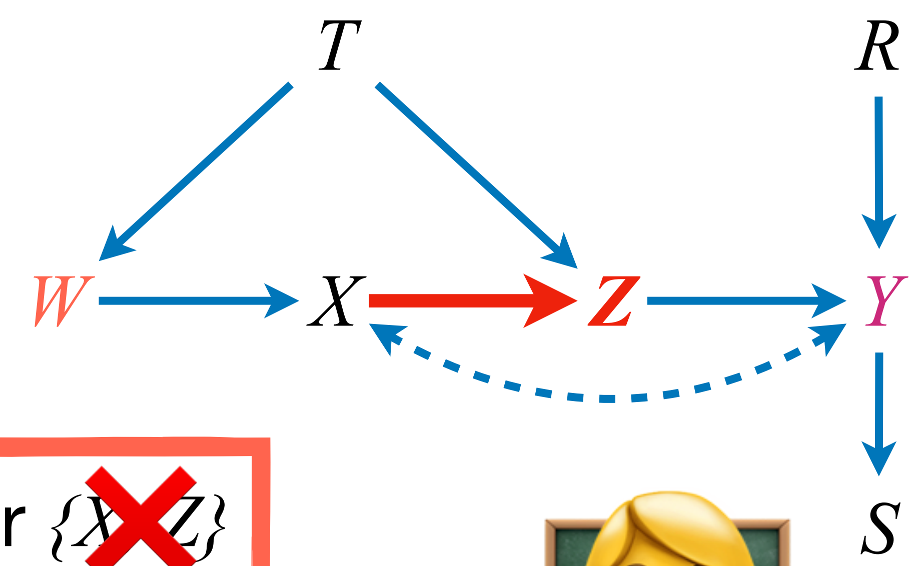
Does $A = \{T, Z\}$ suffice?



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

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: $W \longleftarrow T \longrightarrow Z \longrightarrow Y$

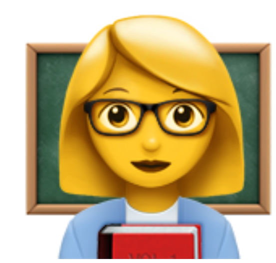
{T} or ~~{Z}~~ or ~~{T, Z}~~

Path 3: $W \longrightarrow X \longleftrightarrow Y$, $X \longrightarrow Z$

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}~~



No such A!

Don't forget the descendants of the colliders!

Food for thought

- Is A independent of D?
- Is A independent of C?
- Is A independent of C given D?
- Is D independent of C given B?