Causal Intutions can help us organize BNs.

Malaria, the flu and a cold all "cause" aches. So use the ordering that places causes before effects. Variables are Malaria (M), Flu (F), Cold (C), Aches (A):

$$P(M,F,C,A) = P(A|M,F,C)\ P(C|M,F)\ P(F|M)\ Pr(M)$$

Each of these disease affects the probability of aches, so the first conditional probability does not change.

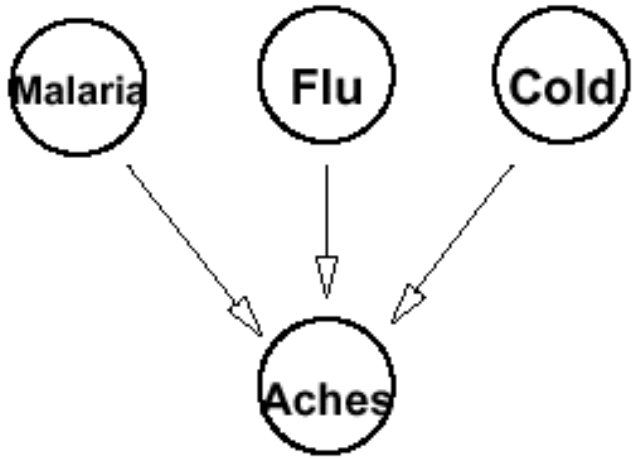
It is however reasonable to assume that these diseases are independent of each other: having or not having one does not change the probability of having the others.

So $P(C|M,F) = P(C)$ and $P(F|M) = P(F)$

Causal Intuitions

This yields a fairly simple Bayes net.

We only need one big CPT, involving the family of "Aches".



Causal Intuitions

Suppose we build the BN for distribution P using the opposite ordering, i.e., we use ordering Aches, Cold, Flu, Malaria

$$P(A,C,F,M) = P(M|A,C,F)\ P(F|A,C)\ P(C|A)\ P(A)$$

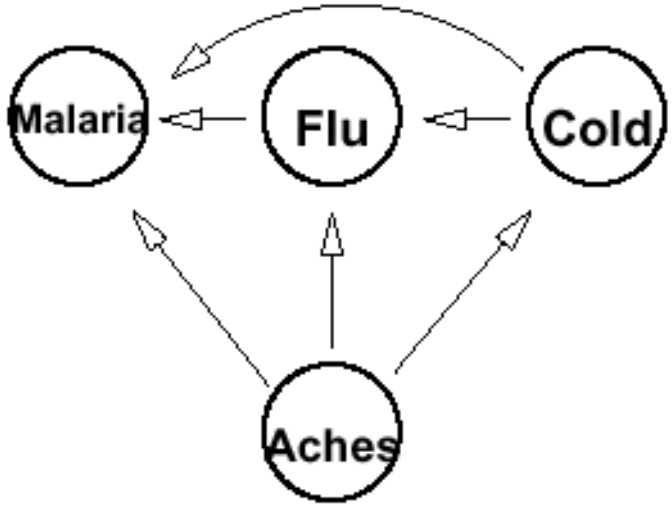
We can't reduce P(M|A,C,F). The probability of Malaria is clearly affected by knowing Aches. What about knowing Aches and Cold, or Aches and Cold and Flu?

Probability of Malaria is affected by both of these additional pieces of knowledge.

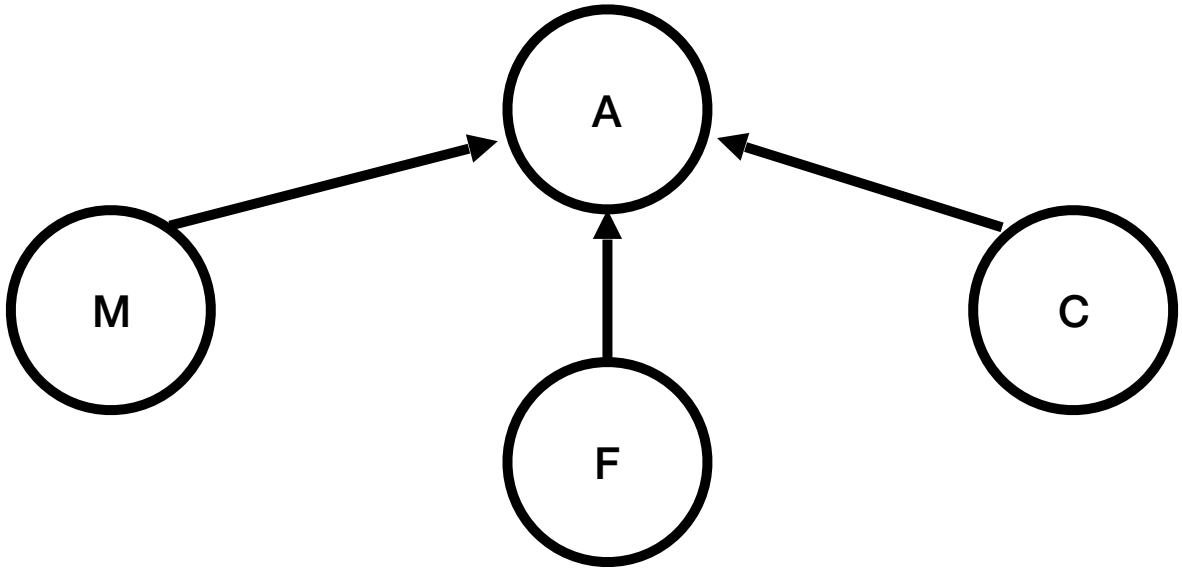
Knowing Cold and Flu lowers the probability that Aches are related to Malaria since they "explain away" the Aches!

Causal Intuitions

We obtain a much more complex Bayes net. In fact, we obtain no savings over explicitly representing the full joint distribution (i.e., representing the probability of every atomic event).



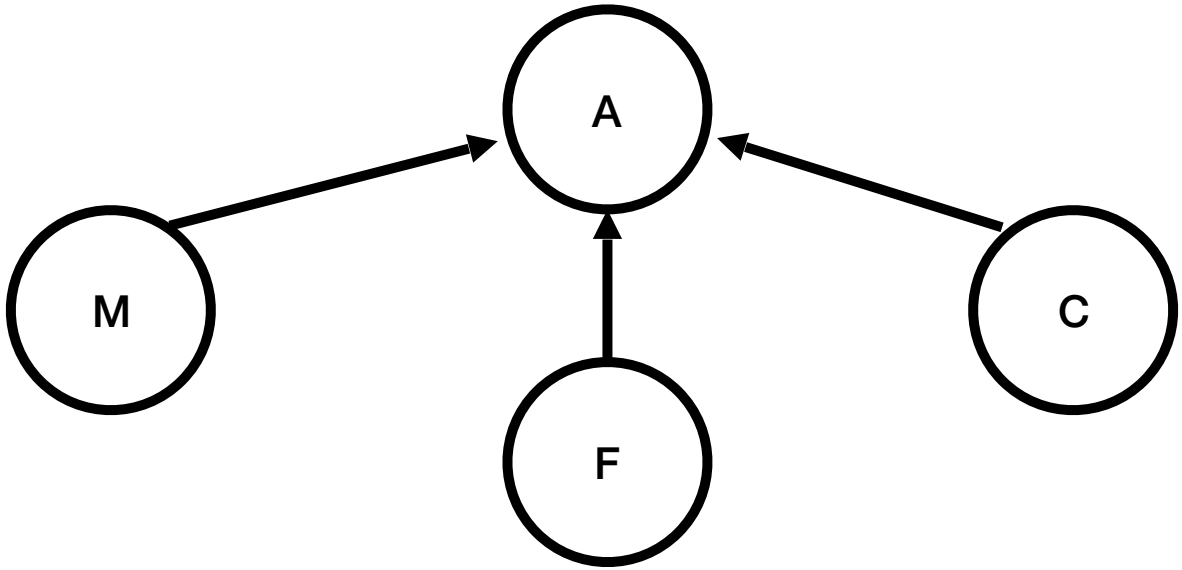
So what have we learned?



Malaria, **Flu** and a **Cold** all *cause* **Aches**! Using causal intuitions led us to an intuitive, simple network that models what we see in data.

But recall, we could have modelled what we see accurately with a *different* network. These *different* networks could all, theoretically, provide a good estimate **P(A|F).**

Causation is *not* Association!



Knowing **P(A|F)** just tells us that Flu is **associated** with Aches. It does **NOT** tell us that Flu **causes** Aches!

To calculate the **causal effect** of Flu on Aches we need access to an **interventional distribution,** which we will call **P(A|do(F=true)).**

Causal Model Example

ID	*T*	*Y*
1	A	0
2	B	1
3	A	0
4	B	1
5	B	0
6	B	1
7	A	1
8	B	1
9	A	0
10	A	0

Let's look at the (synthetic) data summary in the table above. Imagine you have access to the raw data related to a (fictional) COVID-27 pandemic. Assume there are two treatments, A and B, and two possible disease outcomes (Y = 1 means mortality/Y = 0 means survival).

Now you're sick. What do you take … Treatment A or Treatment B?

Causation informed decision

What we want is the relative **causal effect** of the treatments. More specifically, we want to know:

P(Y|do(T=A)) - P(Y|do(T=B))

This represents:
the possibility of mortality when we take treatment A
the possibility of mortality when we take treatment B
everything else must remain equal

But the data we have in hand only lets us calculate:

P(Y|T=A) - P(Y|T=B)

So how to proceed?

Causal Model: New Notation

Let *Ya* be the outcome that we would observe if we **potentially** received treatment *a (i.e. if we do(T=A)).* In this case we could call:

Y_a the outcome under treatment A
Y_b the outcome under treatment B

Y_a and Y_b are referred to as **potential outcomes**

Ideally **and very unrealistically** we could observe **both potential outcomes.**

We will say the treatment T has a **causal effect** on Y if the potential outcomes Y_a and Y_b differ. This causal effect may depend on the individual:

for me, the treatments may have different effects: $\boldsymbol{Y_a \mathrel{!}= Y_b}$
for you, the treatments may have the same effect: $\boldsymbol{Y_a = Y_b}$

Causal Model: New Notation

Say I take treatment A (T = A). Thus, for **me**:

Y_a is observed and equal to the factual outcome *Y. This will be observed in data.*
Yb_0 is unobserved, or **counterfactual**. *This will not be observed in data.*

Let's say you take treatment B (T = B). Thus, for **you**:

Y_b is observed and equal to the factual outcome *Y. This will be observed in data.*
Y_a is unobserved, or **counterfactual**. *This will not be observed in data.*

Note that it is very difficult to say definitively whether treatment A yields better outcomes than treatment B for a specific person **because we cannot observe the outcomes associated with both treatments simultaneously!!**

Causal Model: New Notation

It is easier to justify causal claims when you are looking at data from groups.

e.g. 'if everybody would take Treatment A instead of Treatment B, then the mortality rates would decrease by 15%'

We say treatment A has an average or population level **causal effect** on Y if:

$$\mathbf{P(Y_a = 1) \; != \; P(Y_b = 1)}$$

We say treatment A has no average or population level **causal effect** on Y if:

$$\mathbf{P(Y_a = 1) \; = \; P(Y_b = 1)}$$

Example 1: Ideal Data

ID	Y_a	Y_b
1	0	0
2	1	0
3	0	0
4	1	1
5	0	0
6	1	1
7	1	1
8	1	1
9	0	0
10	1	0

Imagine we were able to observe potential outcomes for 10 individuals both when taking treatment A and treatment B. Let's then calculate the **causal treatment effect**, i.e.:

$$\mathbf{P(Y_a = 1) - P(Y_b = 1)}$$

Example 2: Observed Data

ID	T	Y	Y_b	Y_a
1	A	0	?	0
2	B	1	1	?
3	A	0	?	0
4	B	1	1	?
5	B	0	0	?
6	B	1	1	?
7	A	1	?	1
8	B	1	1	?
9	A	0	?	0
10	A	0	?	0

But in observational studies, we can’t access all potential outcomes!

What if we could?

ID	A	Y	Y_1	Y_0
1	0	0	0	0
2	0	0	0	0
3	0	1	0	1
4	0	1	0	1
5	0	1	0	1
6	0	1	1	1
7	1	0	0	1
8	1	0	0	1
9	1	0	0	0
10	1	1	1	0
11	1	1	1	1
12	1	1	1	1

Table 1: Potential outcome data

Do we get the same **causal treatment effect** if we calculate **$P(Y_0 = 1|A=1) - P(Y_1 = 1|A=0)$?**

Example 2: Observed Data

It is very possible to have situations where

$P(Y_a = 1|T=B) \; != P(Y_a = 1)$ and $P(Y_b = 1|T=A) \; != P(Y_b = 1)$

The causal effect we calculate with observations is therefore **NOT** equal to the true causal effect. Why?

Because the distributions of outcomes carry information about group membership (and are therefore *not* **exchangeable)**, i.e.

Y_a and Y_b are not independent of T

Observed outcomes may **confounded** by properties of group membership.

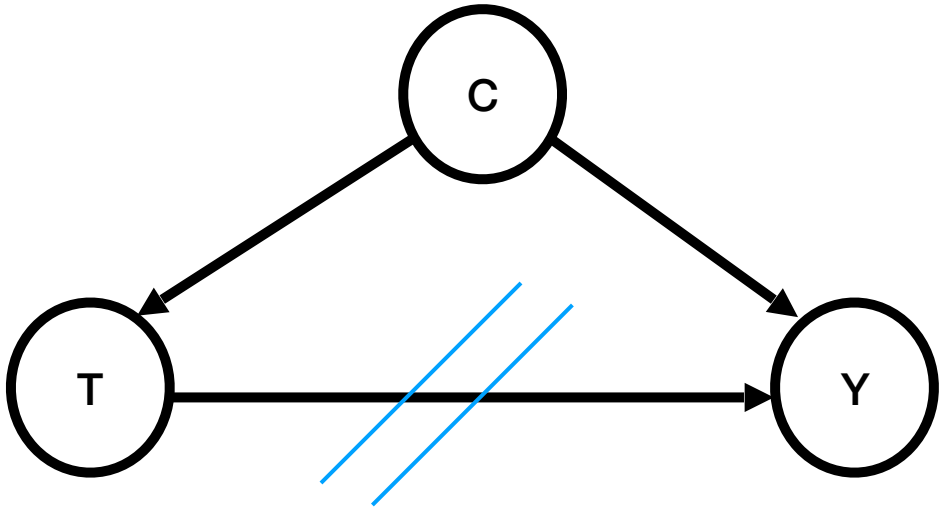
The take home message: ***Association != Causation***

What causes non-exchangeability?

Suppose there is a covariate (e.g. C) that effects **both T and Y.**

T will then be associated with Y, even if there is no causal effect!

This is an example of a **confounding** variable**.**



Example 3: Potential Confounders

Treatment	Condition: Mild	Condition: Severe	Condition: Total
A	15% (210/1400)	30% (30/100)	**16%** (240/1500)
B	**10%** (5/50)	**20%** (100/500)	19% (105/550)

Table 1.1: Simpson's paradox in COVID-27 data. The percentages denote the mortality rates in each of the groups. Lower is better. The numbers in parentheses are the corresponding counts. This apparent paradox stems from the interpretation that treatment A looks better when examining the whole population, but treatment B looks better in all subpopulations.

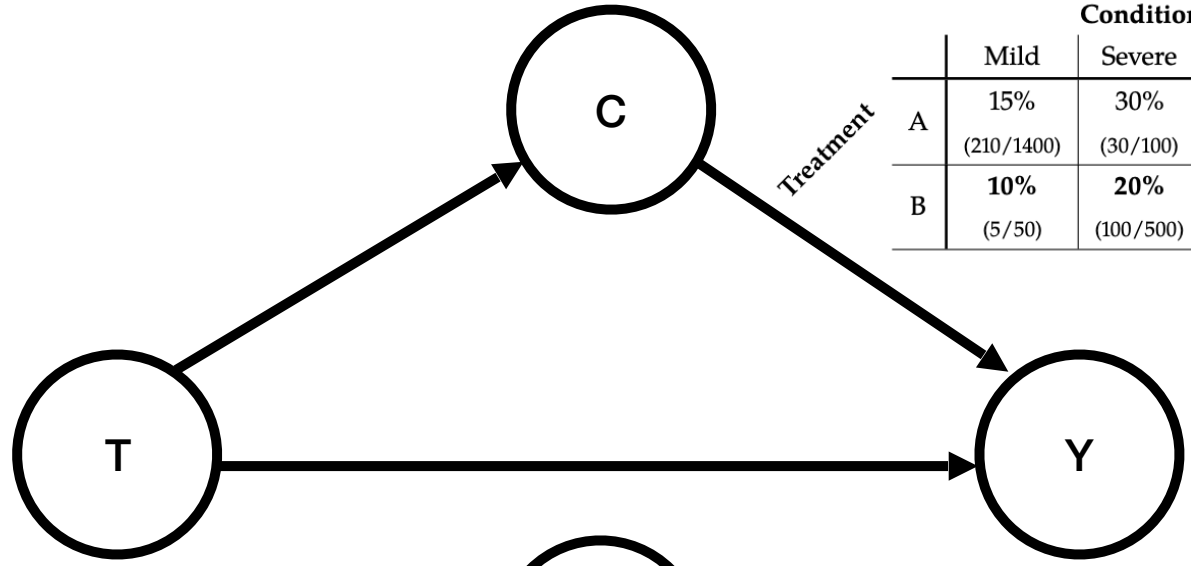
Let's look again at a (synthetic) data summary in the table above. Again, there are two treatments, A and B and two possible disease outcomes (mortality/survival). We have added a third variable (condition) which can be mild and severe.

Now you're sick. What do you take … Treatment A or Treatment B?

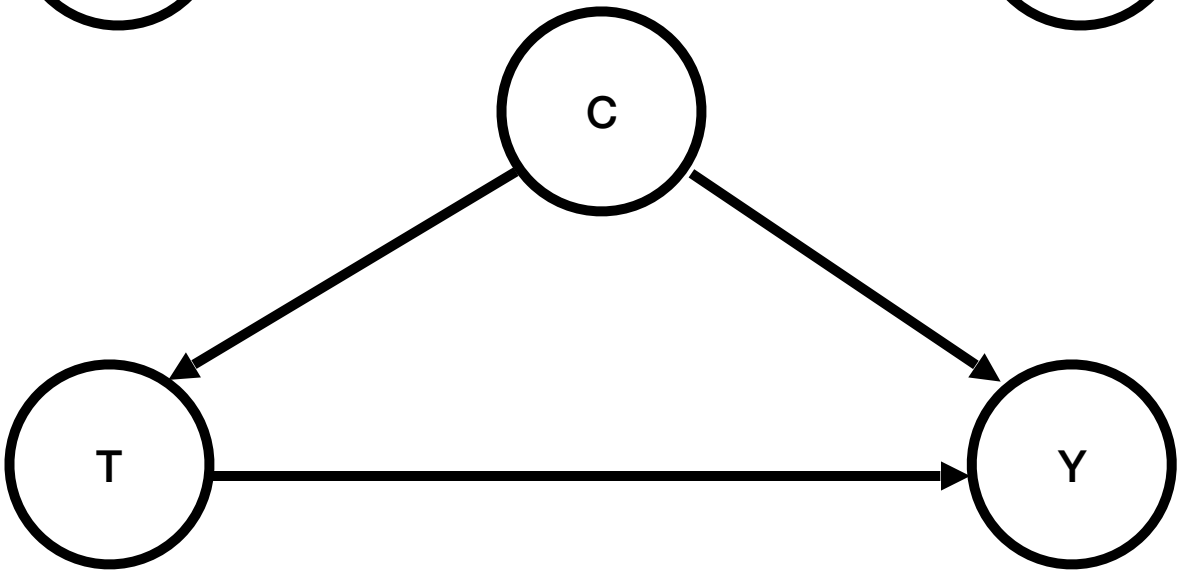
Your decision depends on your story!

Treatment	Condition: Mild	Condition: Severe	Condition: Total
A	15% (210/1400)	30% (30/100)	**16%** (240/1500)
B	**10%** (5/50)	**20%** (100/500)	19% (105/550)

Table 1.1: Simpson's paradox in COVID-27 data. The percentages denote the mortality rates in each of the groups. Lower is better. The numbers in parentheses are the corresponding counts. This apparent paradox stems from the interpretation that treatment A looks better when examining the whole population, but treatment B looks better in all subpopulations.



Story 1: The treatment **causes** the severity. Get treatment B, you get severe COVID. Get treatment A, you get mild COVID. Here C is a **mediating** variable.



Story 2: The disease severity **causes** the treatment. People with severe COVID get treatment B. People with mild COVID get treatment A. Here C is a **confounding** variable.

To know which model correctly explains our observations, **we must depend on experts for help.**

Adjusting for Confounders

There are several ways to adjust for confounders.

Stratification

Matching

Propensity Scoring

Inverse Probability Weighting

and ….

Adjusting for Confounders: Standardization

We can conditionally adjust for confounders if we can locate some set of variables W such that:

Y_a and Y_b are independent of T given W

Exchangeability can be achieved by adjustments, but we can also **destroy exchangeability** if we adjust for the wrong variables! *Think about what you know about d-separation here! We will circle back to this issue.*

To locate the right adjustment set W for a given problem, **we cannot get by without the help of experts.**

Example 3: Potential Confounders

Treatment	Condition: Mild	Condition: Severe	Condition: Total
A	15% (210/1400)	30% (30/100)	**16%** (240/1500)
B	**10%** (5/50)	**20%** (100/500)	19% (105/550)

Table 1.1: Simpson's paradox in COVID-27 data. The percentages denote the mortality rates in each of the groups. Lower is better. The numbers in parentheses are the corresponding counts. This apparent paradox stems from the interpretation that treatment A looks better when examining the whole population, but treatment B looks better in all subpopulations.

Assume experts have told us disease condition is a confounding variable in our network. Let's then calculate these conditional **causal treatment effects**, i.e.:

$$P(Y_a = 1|C=Mild) - P(Y_b = 1|C=Mild)$$

and

$$P(Y_a = 1|C=Severe) - P(Y_b = 1|C=Severe)$$

Example 3: Potential Confounders

If we have conditional exchangeability, given *C*, we can calculate the average treatment effect $P(Y_a = 1)$ through **standardization:**

$$P(Y_a = 1) = \sum_c P(Y_a = 1 \mid C = c) * P(C = c)$$

$$= \sum_c P(Y = 1 \mid T = a, C = c) * P(C = c)$$

Example 3: Potential Confounders

Treatment	Condition: Mild	Condition: Severe	Condition: Total
A	15% (210/1400)	30% (30/100)	**16%** (240/1500)
B	**10%** (5/50)	**20%** (100/500)	19% (105/550)

Table 1.1: Simpson's paradox in COVID-27 data. The percentages denote the mortality rates in each of the groups. Lower is better. The numbers in parentheses are the corresponding counts. This apparent paradox stems from the interpretation that treatment A looks better when examining the whole population, but treatment B looks better in all subpopulations.

Let's assume we have conditional exchangeability given **condition**. Use this information to calculate the adjusted **causal treatment effect**, i.e.:

$$P(Y_a = 1) - P(Y_b = 1)$$

Review

Y_a = P(Y|do(T=a)) represents the effect on Y if we intervene to set T to a. We need this value to estimate the **causal effect** of setting T to a.

Y_a = P(Y|do(T=a)) *is not the same as* ***P(Y|T=a)***

However, if we have **exchangeability in observed data**, we can use observed associations to calculate causal effects.

Exchangeability requires:

$$(Y_a, Y_b) \perp\!\!\!\perp T$$

*We typically **don't** have exchangeability in observational data.*

If we make sufficient confounder adjustments to achieve exchangeability (i.e. by standardization) we can calculate causal effects. But knowing our adjustments are sufficient requires expert knowledge to determine.

It is moreover possible to **destroy exchangeability** if we adjust for the wrong variables!

Why might this be true?

Parting Thoughts

Why does this matter to practitioners of AI/ML?

1. We make agents that **do things!** Understanding the consequences of our actions helps our agents make robust, meaningful decisions that generalize.

2. We **do things**, too! If we know **why** things happen causally, we can take actions to improve our individual and collective outcomes. We might stop smoking, log off instagram, take a walk in the park! Causation is an integral part of reasoning in many disciplines like medicine for exactly this reason.

Thanks

Several examples in slides are drawn from Arvid Sjolander, Dept. of Medical Epidemiology and Biostatistics, Karolinska Institutet (Sweden)