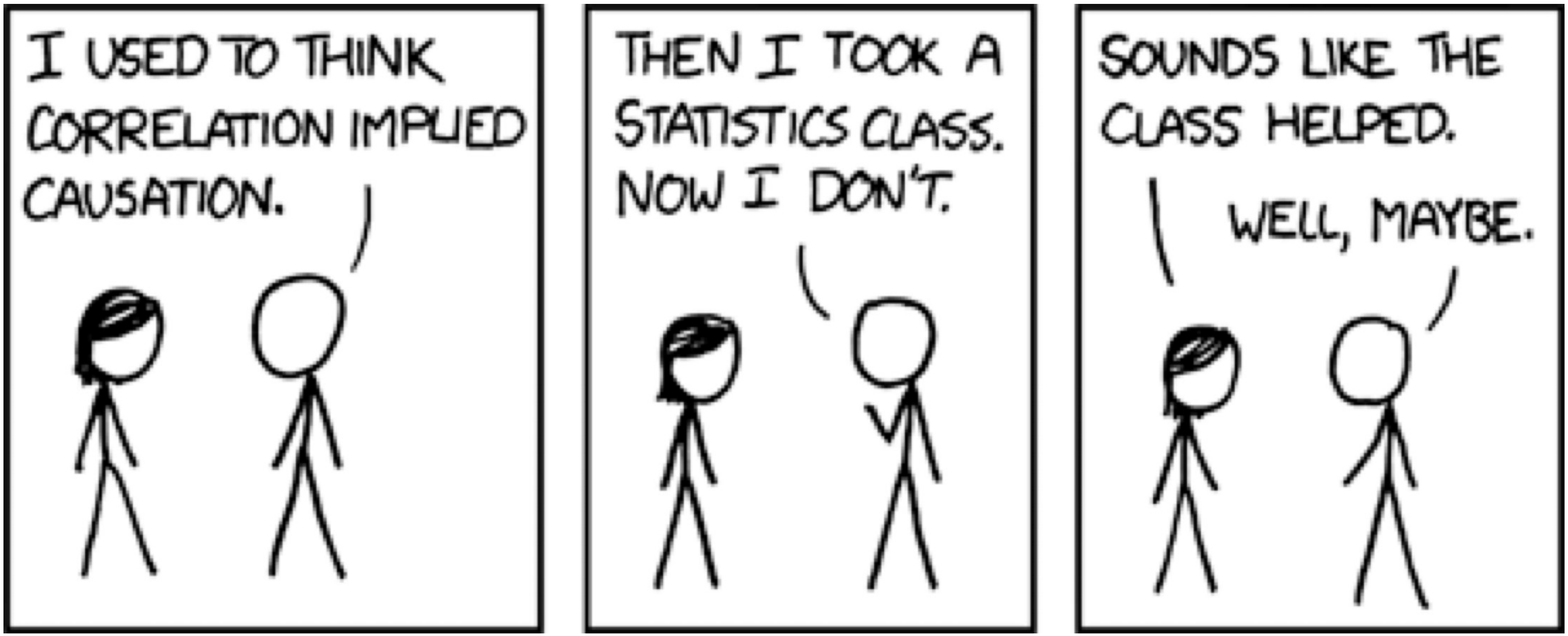
Causal Graphical Models



Zach Wood-Doughty and Bryan Pardo

CS 349 Fall 2021

Some slides taken from Ilya Shpitser

If you're interested in this material…

ACADEMICS / COURSES / DESCRIPTIONS

COMP_SCI 396: Modeling Relationships with Causal Inference

Quarter Offered

Winter : 5-6:20 MW ; Wood-Doughty

Prerequisites

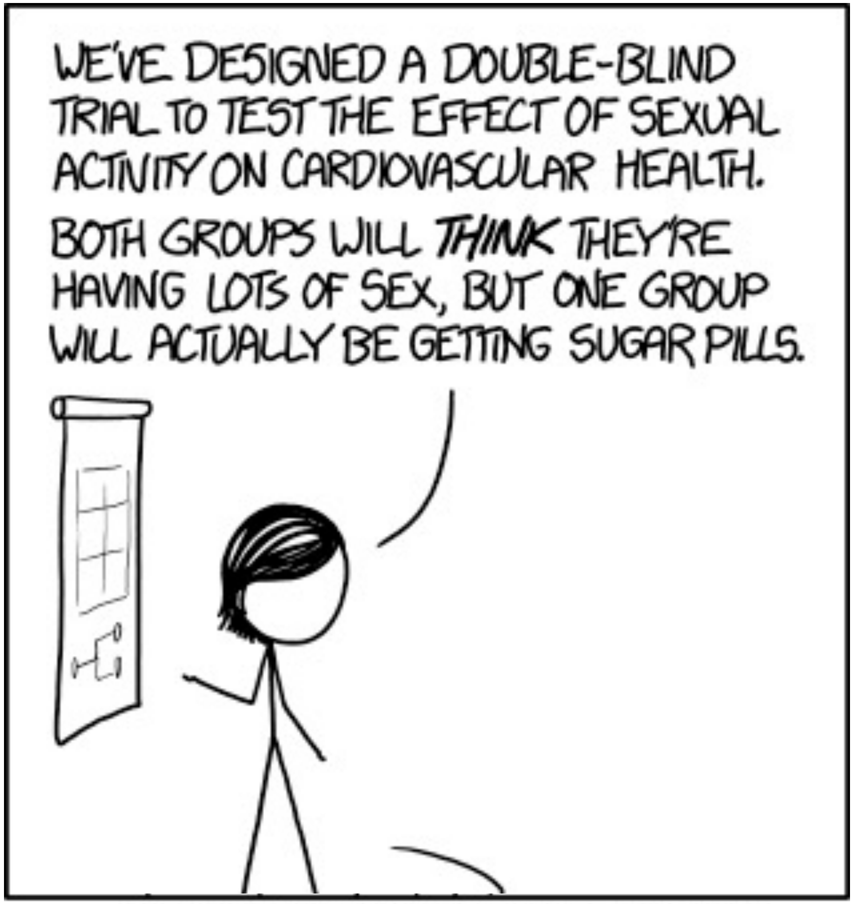
Permission of Instructor

Correlation doesn't imply causation, but…

- Carefully analyzing correlations is often our best or only approach to inferring causation

- Randomized experiments are impossible or unethical in many domains

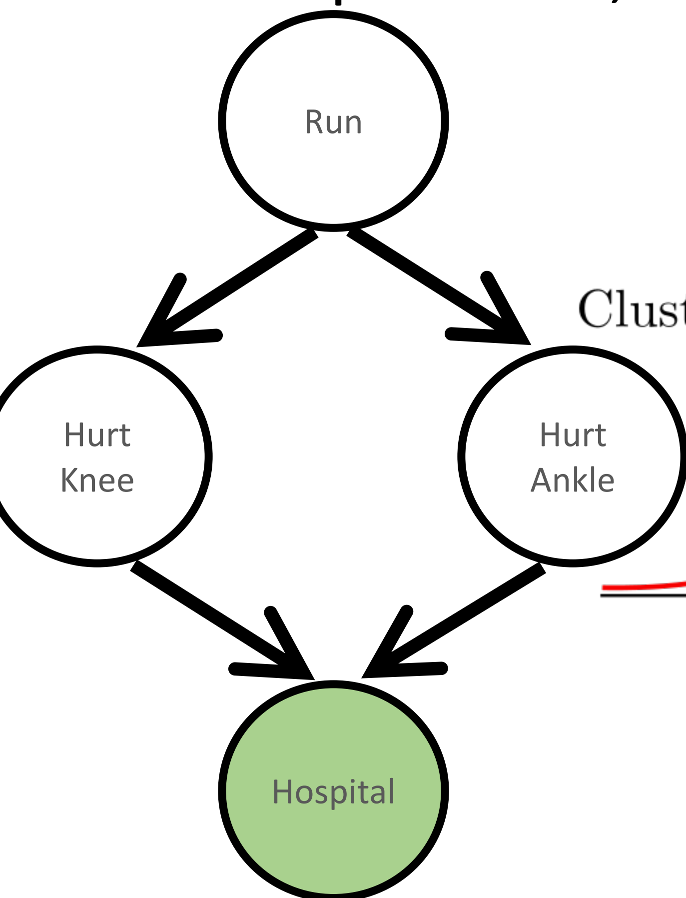
- There are many methods that can enable valid causal inferences from observational data

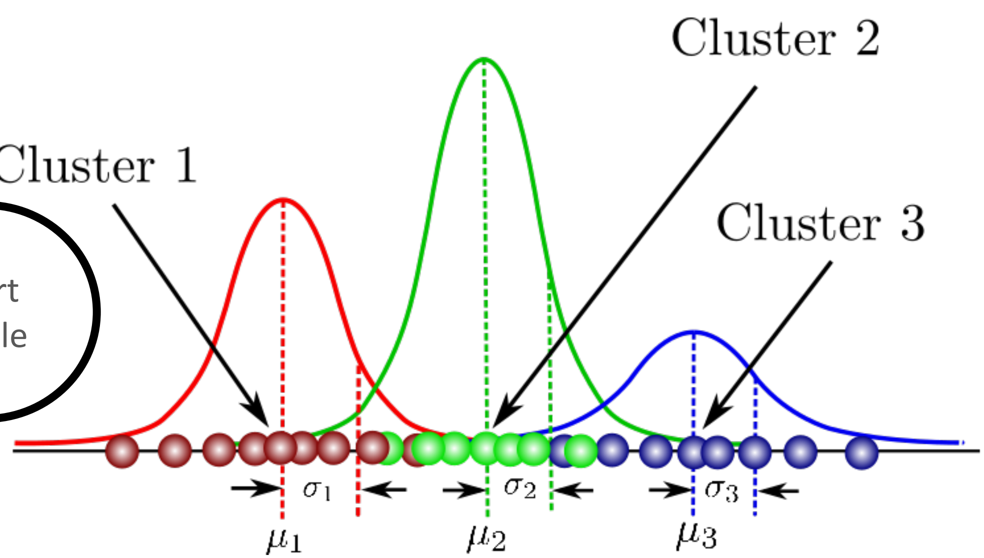


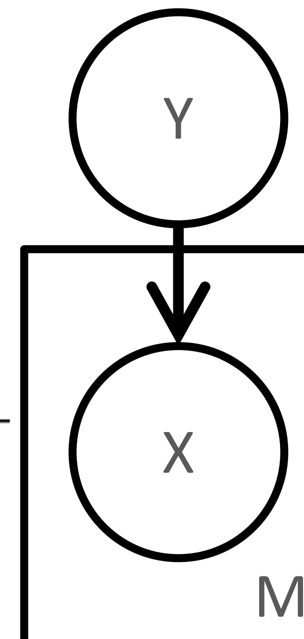
THE LIMITATIONS OF BLIND TRIALS

Probabilistic versus causal models

- Many models we've considered have a *probabilistic* interpretation, but not necessarily a *causal* one





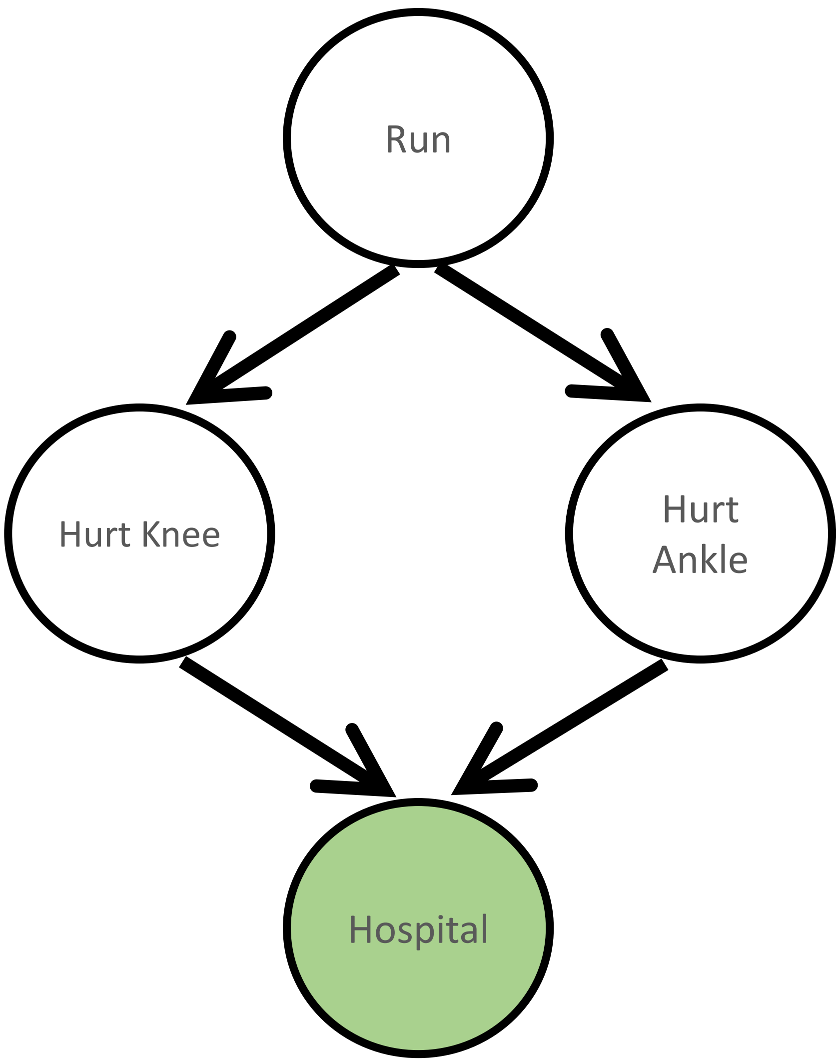


How do we make our model *causal*?

- Causal assumptions are what separates whether we believe a graphical model:
    1. Helps understand the data and predict the future, or
    2. Accurately represents the real-world phenomena that generated the data

- If these assumptions are correct, we can make valid causal claims. If they're wrong, our claims may be arbitrarily wrong!

A causal generative story



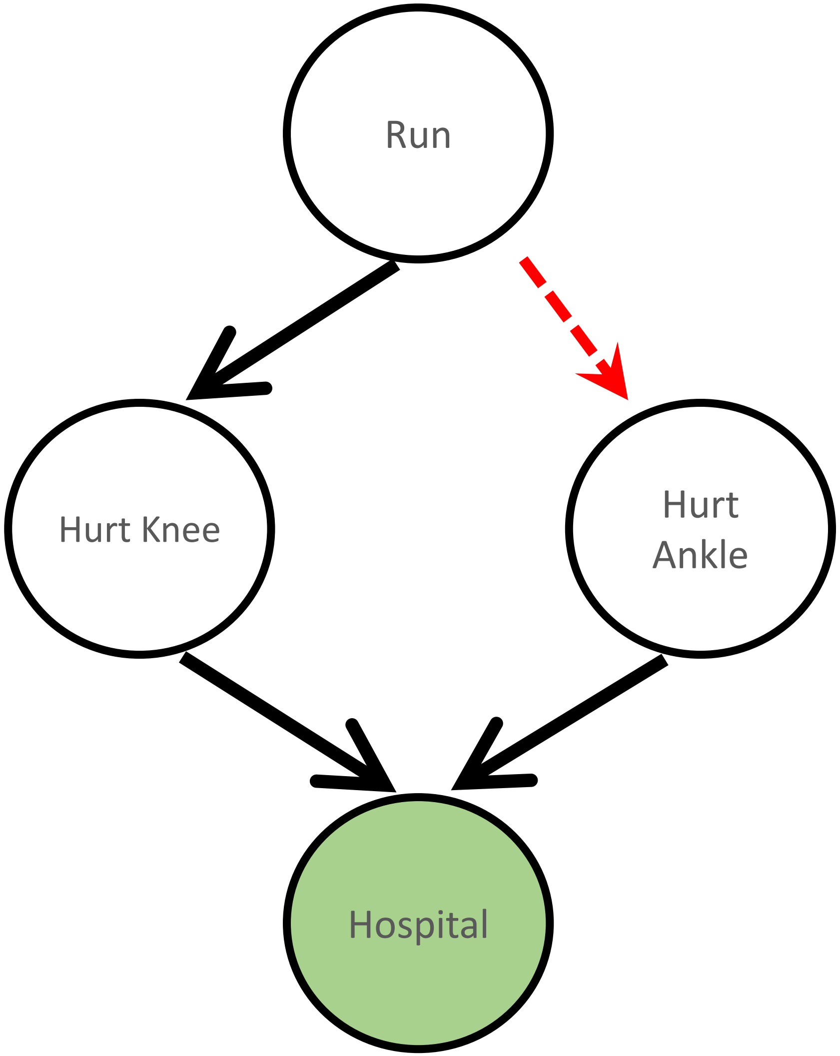
P(Run=T)
.5

Run	P(Knee=T)
F	0.01
T	0.1

Run	P(Ankle=T)
F	0.05
T	0.2

Ankle	Knee	P(Hospital=T)
F	F	0
T	F	0.3
F	T	.9
T	T	.99

A causal generative story



Ankle	Knee	P(Hospital=T)
F	F	0
T	F	0.3
F	T	.9
T	T	.99

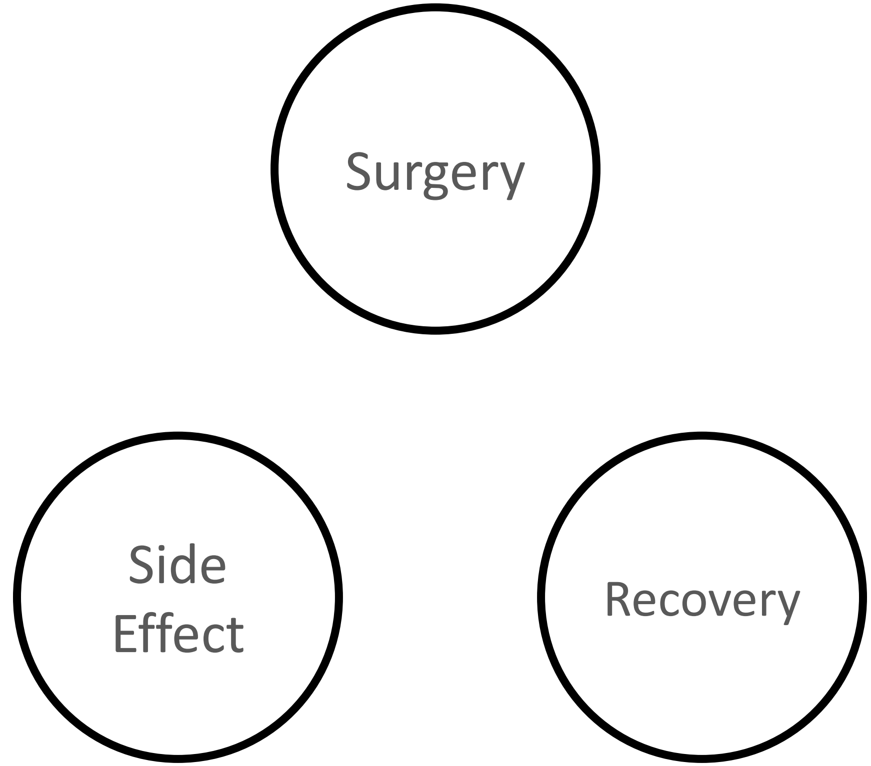
Why is causal inference hard?

% who recover from sports injury	**Surgery A**	**Surgery B**
Without side-effect	93% (81/87)	87% (234/270)
With side-effect	73% (192/263)	69% (55/80)
Both	78% (273/350)	83% (289/350)

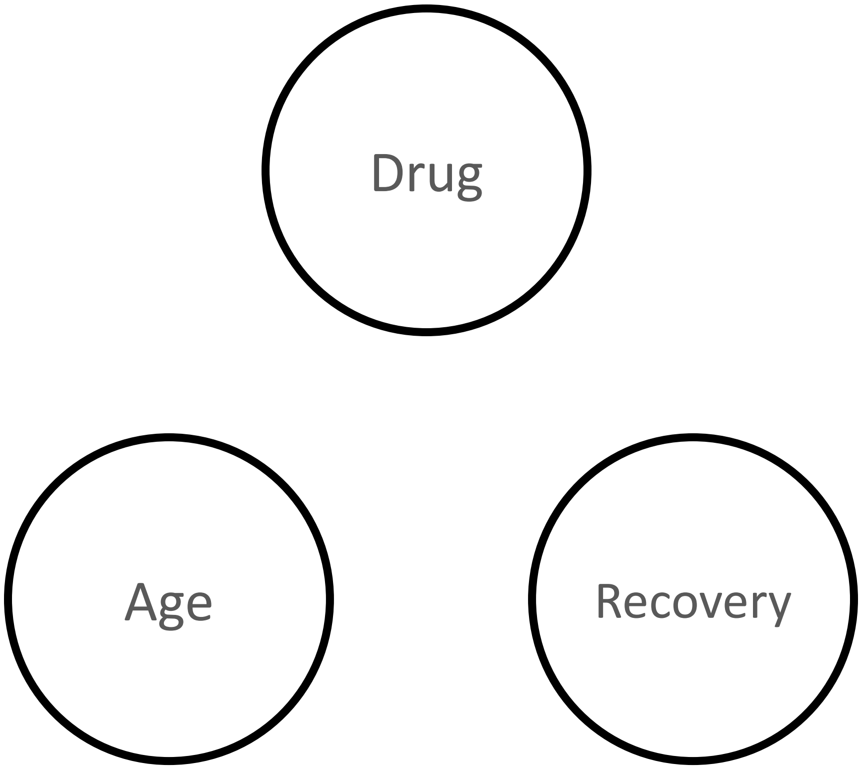
% who recover from migraines	**Drug C**	**Drug D**
Younger patients	93% (81/87)	87% (234/270)
Older patients	73% (192/263)	69% (55/80)
Both	78% (273/350)	83% (289/350)

Causal assumptions

% who recover from sports injury	**Surgery A**	**Surgery B**
Without side-effect	93% (81/87)	87% (234/270)
With side-effect	73% (192/263)	69% (55/80)
Both	78% (273/350)	83% (289/350)



Causal assumptions



% who recover from migraines	**Drug C**	**Drug D**
Younger patients	93% (81/87)	87% (234/270)
Older patients	73% (192/263)	69% (55/80)
Both	78% (273/350)	83% (289/350)

Why is causal inference hard?

% who recover from sports injury	**Surgery A**	**Surgery B**
Without side-effect	93% (81/87)	87% (234/270)
With side-effect	73% (192/263)	69% (55/80)
Both	78% (273/350)	83% (289/350)

% who recover from migraines	**Drug C**	**Drug D**
Younger patients	93% (81/87)	87% (234/270)
Older patients	73% (192/263)	69% (55/80)
Both	78% (273/350)	83% (289/350)

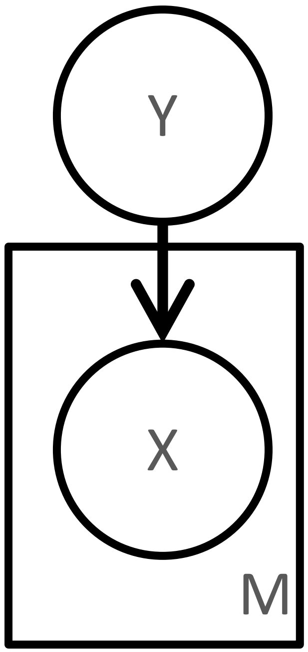
Why is causal inference hard?

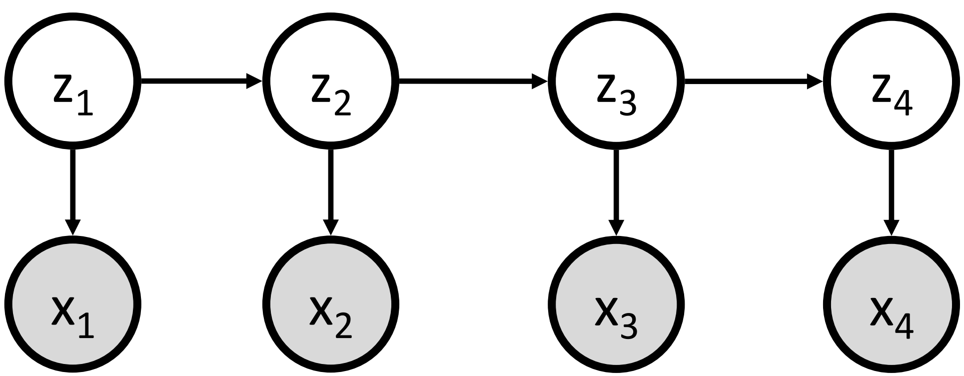
Leaderboard

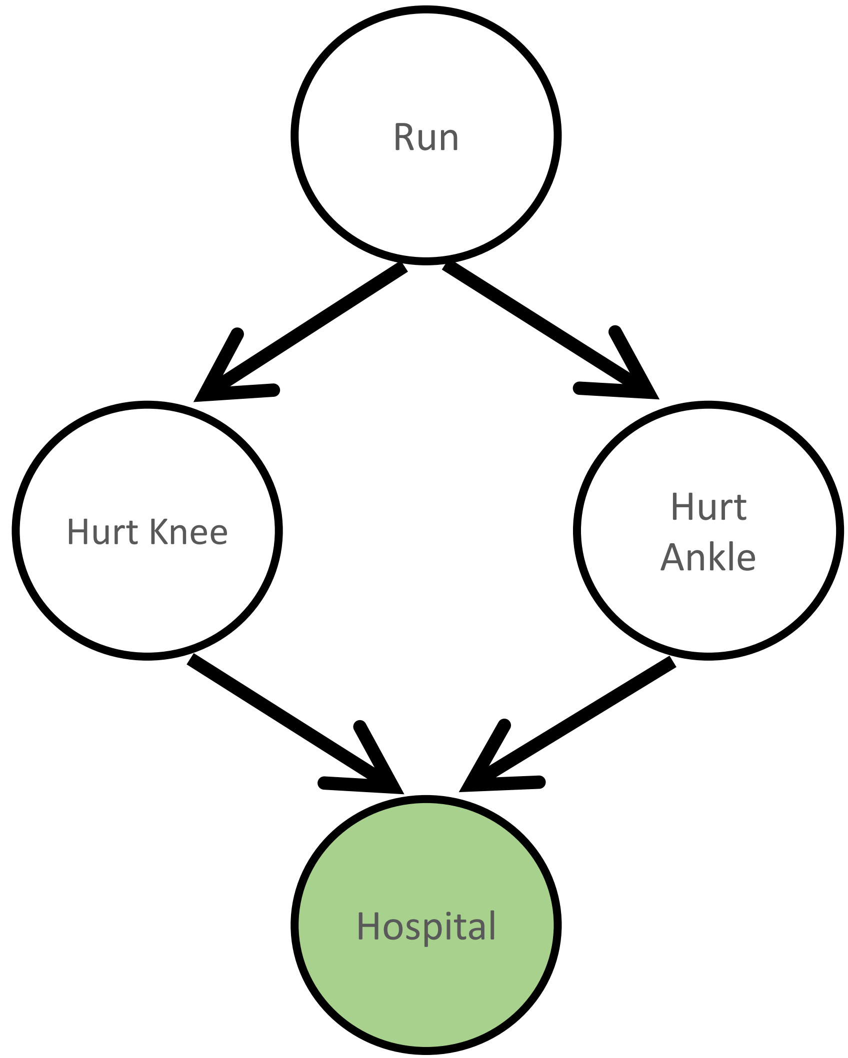
SQuAD2.0 tests the ability of a system to not only answer reading comprehension questions, but also abstain when presented with a question that cannot be answered based on the provided paragraph.

Rank	Model	EM	F1
	Human Performance *Stanford University* (Rajpurkar & Jia et al. '18)	86.831	89.452
1 Apr 06, 2020	SA-Net on Albert (ensemble) *QIANXIN*	**90.724**	**93.011**
2 May 05, 2020	SA-Net-V2 (ensemble) *QIANXIN*	90.679	92.948

Assumptions in ML vs causal inference







Aspirin and CVD, 2002, 2016, and 2021

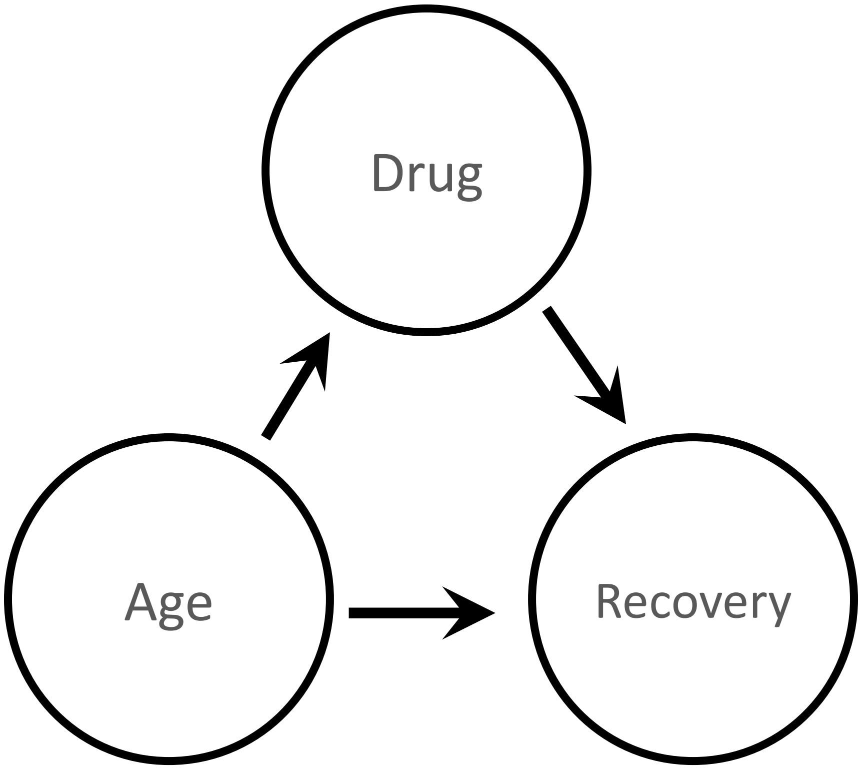
Population	Recommendation	Grade
Adults with are at increased risk for coronary heart disease (CHD)	The U.S. Preventive Services Task Force (USPSTF) strongly recommends that clinicians discuss aspirin chemoprevention with adults who are at increased risk for coronary heart disease (CHD) (go to Clinical Considerations). Discussions with patients should address both the potential benefits and harms of aspirin therapy	A

Adults aged 60 to 69 years with a 10% or greater 10-year CVD risk	The decision to initiate low-dose aspirin use for the primary prevention of CVD and CRC in adults aged 60 to 69 years who have a 10% or greater 10-year CVD risk should be an individual one. Persons who	C

Recommendation Summary

Population	Recommendation	Grade
Adults ages 40 to 59 years with a 10% or greater 10-year cardiovascular disease (CVD) risk	The decision to initiate low-dose aspirin use for the primary prevention of CVD in adults ages 40 to 59 years who have a 10% or greater 10-year CVD risk should be an individual one. Evidence indicates that the net benefit of aspirin use in this group is small. Persons who are not at increased risk for bleeding and are willing to take low-dose aspirin daily are more likely to benefit.	C
Adults age 60 years or older	The USPSTF recommends against initiating low-dose aspirin use for the primary prevention of CVD in adults age 60 years or older.	D

Causal assumptions

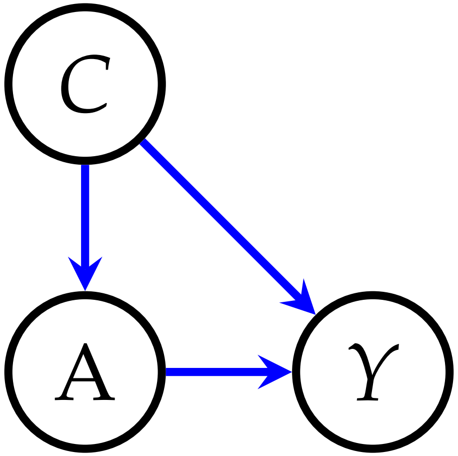


% who recover from migraines	**Drug C**	**Drug D**
Younger patients	93% (81/87)	87% (234/270)
Older patients	73% (192/263)	69% (55/80)
Both	78% (273/350)	83% (289/350)

Counterfactual random variables

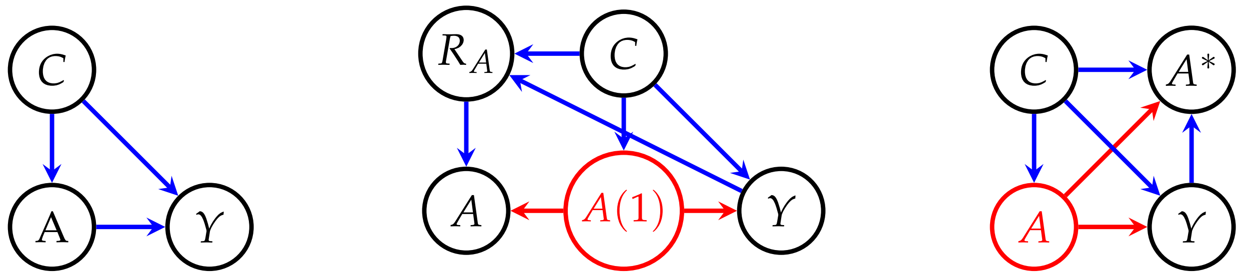
ID	Age	Drug	Recover (C)	Recover (D)
1	Old	C	Yes	*No*
2	Young	C	Yes	*No*
3	Young	C	No	*No*
4	Young	D	*Yes*	Yes
5	Old	D	*No*	No

Derivation of the causal effect



$$
\begin{aligned}
p(Y(a)) &= \sum_C p(Y(a) \mid C) p(C) \\
&= \sum_C p(Y(a) \mid A, C) p(C) \\
&= \sum_C p(Y \mid A, C) p(C)
\end{aligned}
$$

Connections back to machine learning



(a) Simple Confounding **(b)** Missing Data **(c)** Measurement Error

Figure 2-2. DAGs for causal inference. Red variables are unobserved. A is a treatment, Y is an outcome, and C is a confounder.

A^*	C	Y
0	1	0
0	1	1
0	0	1
1	0	1

(c) Measurement Error

A^*	A
1	1
0	1
0	0
1	1

(d) Mismeasurement

Causal inference with ML methods

Zika Virus as a Cause of Neurologic Disorders

Nathalie Broutet, M.D., Ph.D., Fabienne Krauer, M.Sc., Maurane Riesen, M.Sc., Asheena Khalakdina, Ph.D., Maria Almiron, M.Sc., Sylvain Aldighieri, M.D., Marcos Espinal, M.D., Nicola Low, M.D., and Christopher Dye, D.Phil.

Zika Virus and Birth Defects — Reviewing the Evidence for Causality

Sonja A. Rasmussen, M.D., Denise J. Jamieson, M.D., M.P.H., Margaret A. Honein, Ph.D., M.P.H., and Lyle R. Petersen, M.D., M.P.H.

Causal inference with ML methods

cardiothoracic surgery history of present illness seventy two year old **retired pediatric cardiologist** presents with increasing angina and shortness of breath a stress test performed in was positive ejection fraction was past medical history hypertension hypercholesterolemia **cigarette smoking but quit** in the gastrointestinal bleeding in he has never had a stroke tia or claudication

If you're interested in this material...

ACADEMICS / COURSES / DESCRIPTIONS

COMP_SCI 396: Modeling Relationships with Causal Inference

Quarter Offered

Winter : 5-6:20 MW ; Wood-Doughty

Prerequisites

Permission of Instructor