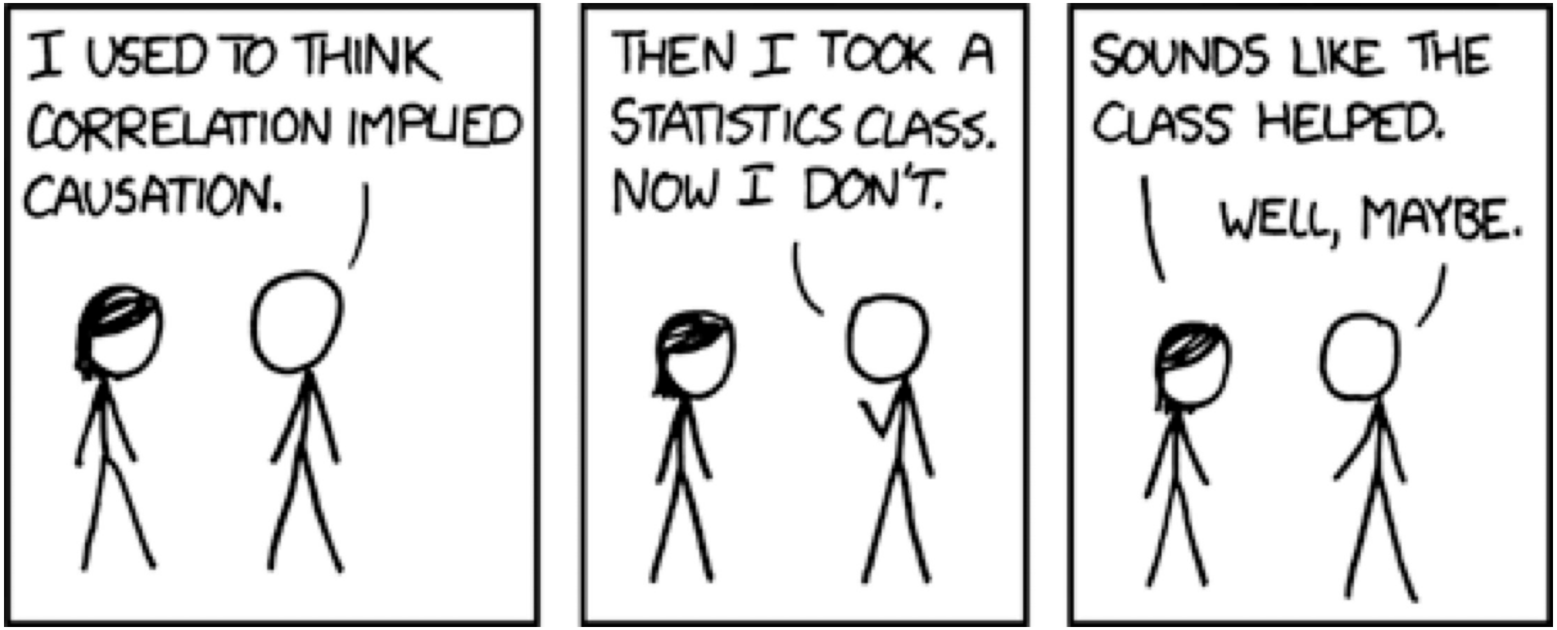
Modeling Relationships with Causal Inference

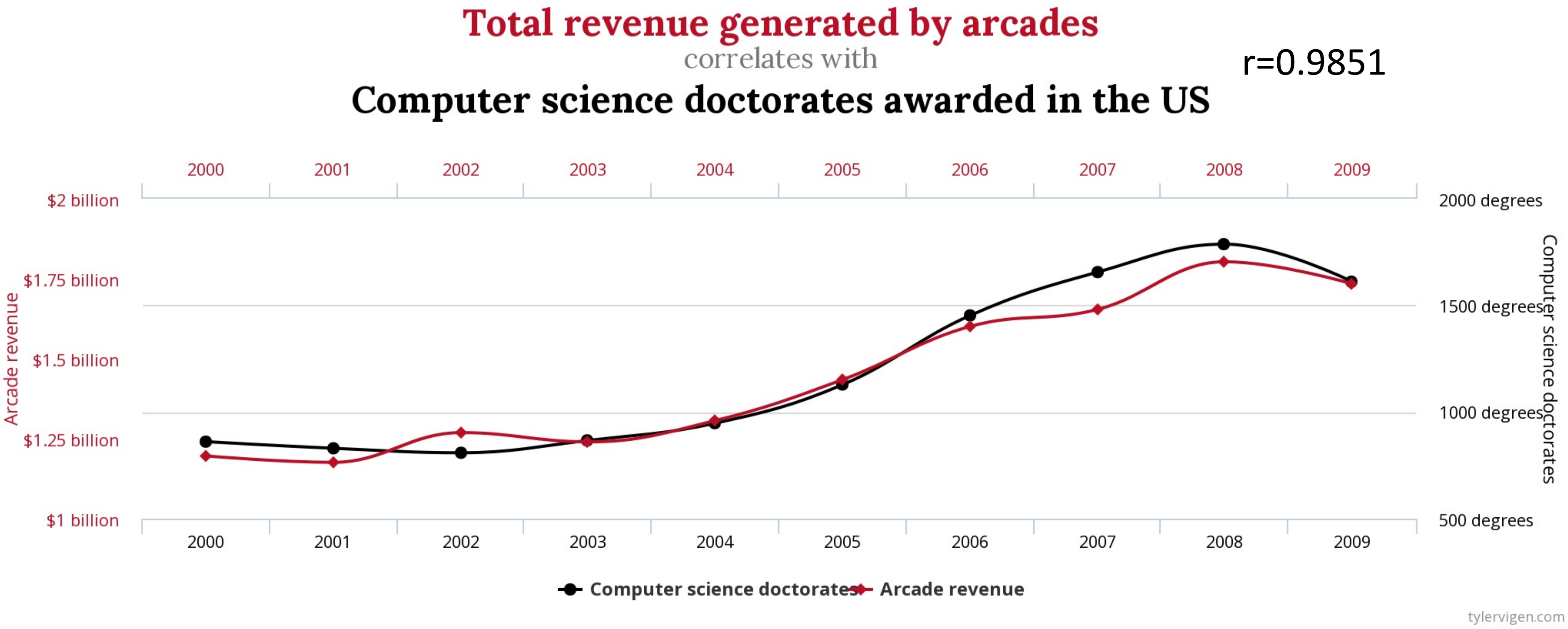

I USED TO THINK CORRELATION IMPLIED CAUSATION.
THEN I TOOK A STATISTICS CLASS. NOW I DON'T.
SOUNDS LIKE THE CLASS HELPED.
WELL, MAYBE.

Zach Wood-Doughty CS 396 Winter 2022

Today

- Motivating the need for causal inference
- Understanding Simpson’s paradox
- Course policies and rough schedule
- Start thinking about the final project

Correlation does not imply causation

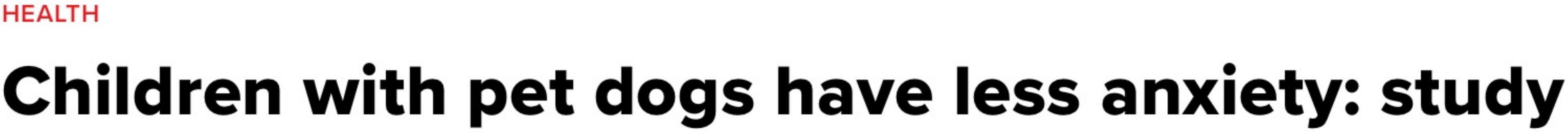


“Did you know that pet ownership has great benefits for your health?”

- Pet owners are less likely to suffer from stress, anxiety, and depression than non-pet owners.
- The NIH found that dog owners who walk their dogs are significantly more likely to meet physical activity guidelines.
- Pet ownership is associated with key indicators of cardiovascular health such as lower blood pressure.

From US Public Health Service flyer, “Pets Promote Health”

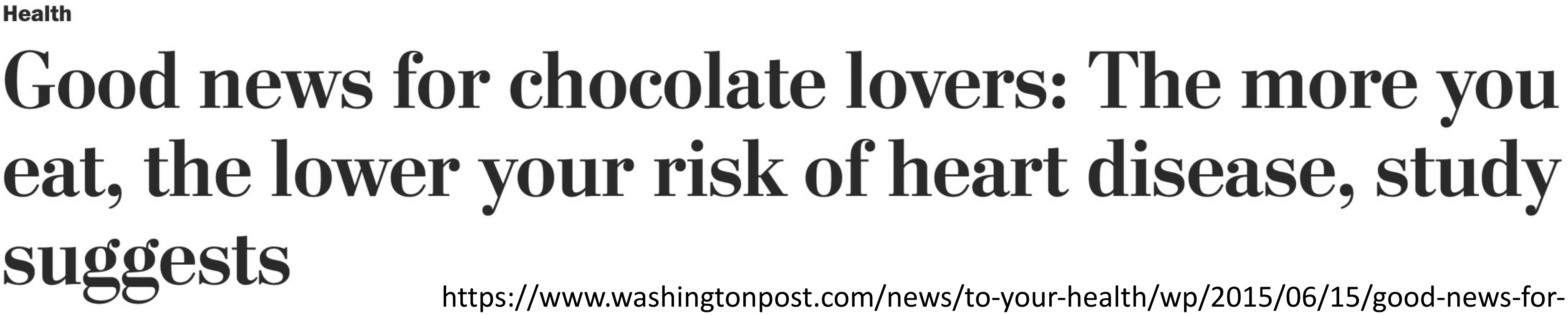
https://www.cdc.gov/pcd/issues/2015/15_0204.htm



HEALTH

Children with pet dogs have less anxiety: study

https://globalnews.ca/news/2367202/children-with-pet-dogs-have-less-anxiety-study/



Health

Good news for chocolate lovers: The more you eat, the lower your risk of heart disease, study suggests

https://www.washingtonpost.com/news/to-your-health/wp/2015/06/15/good-news-for-chocolate-lovers-study-suggests-the-more-you-eat-the-lower-your-risk-of-heart-disease/

Why is causal inference hard?

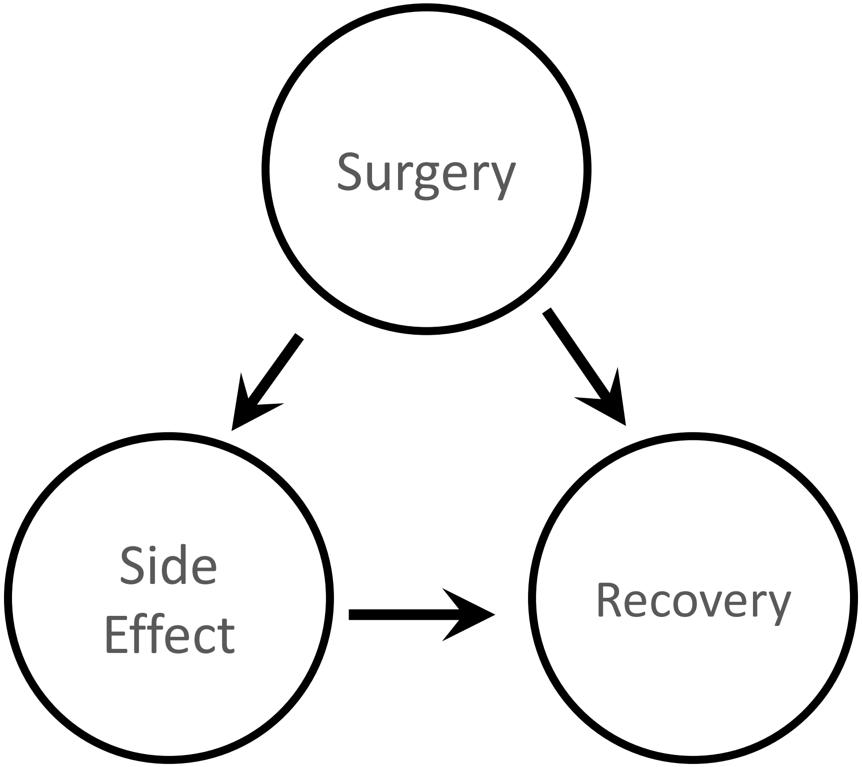
% who recover from sports injury	**Surgery A**	**Surgery B**
Without side-effect	93% (81/87)	87% (234/270)
With side-effect	73% (192/263)	69% (55/80)
Both	78% (273/350)	83% (289/350)

Causal assumptions

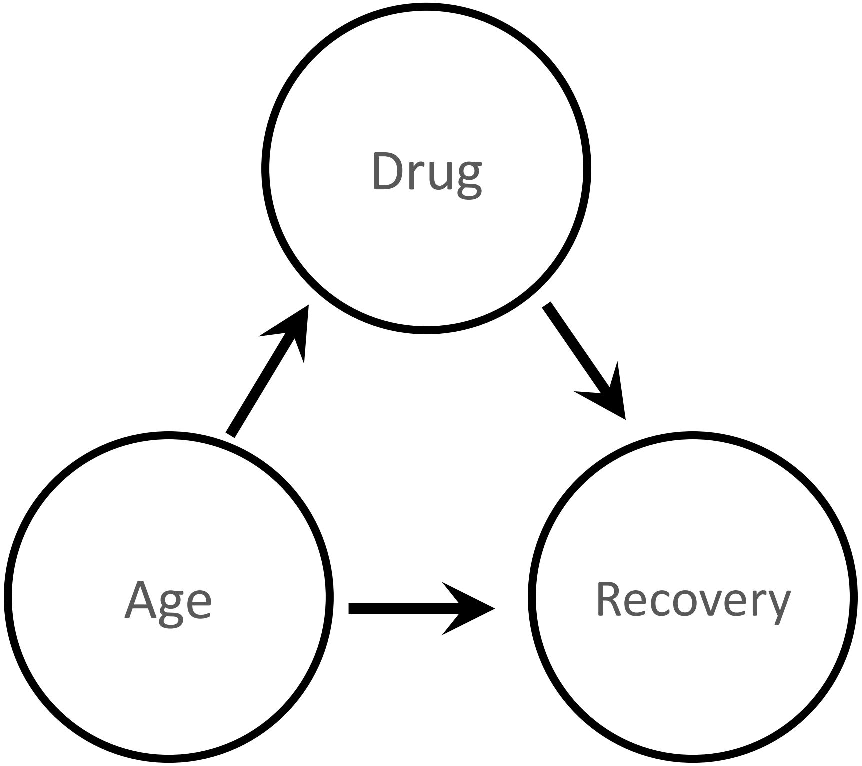
% who recover from migraines	**Drug C**	**Drug D**
Younger patients	93% (81/87)	87% (234/270)
Older patients	73% (192/263)	69% (55/80)
Both	78% (273/350)	83% (289/350)

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



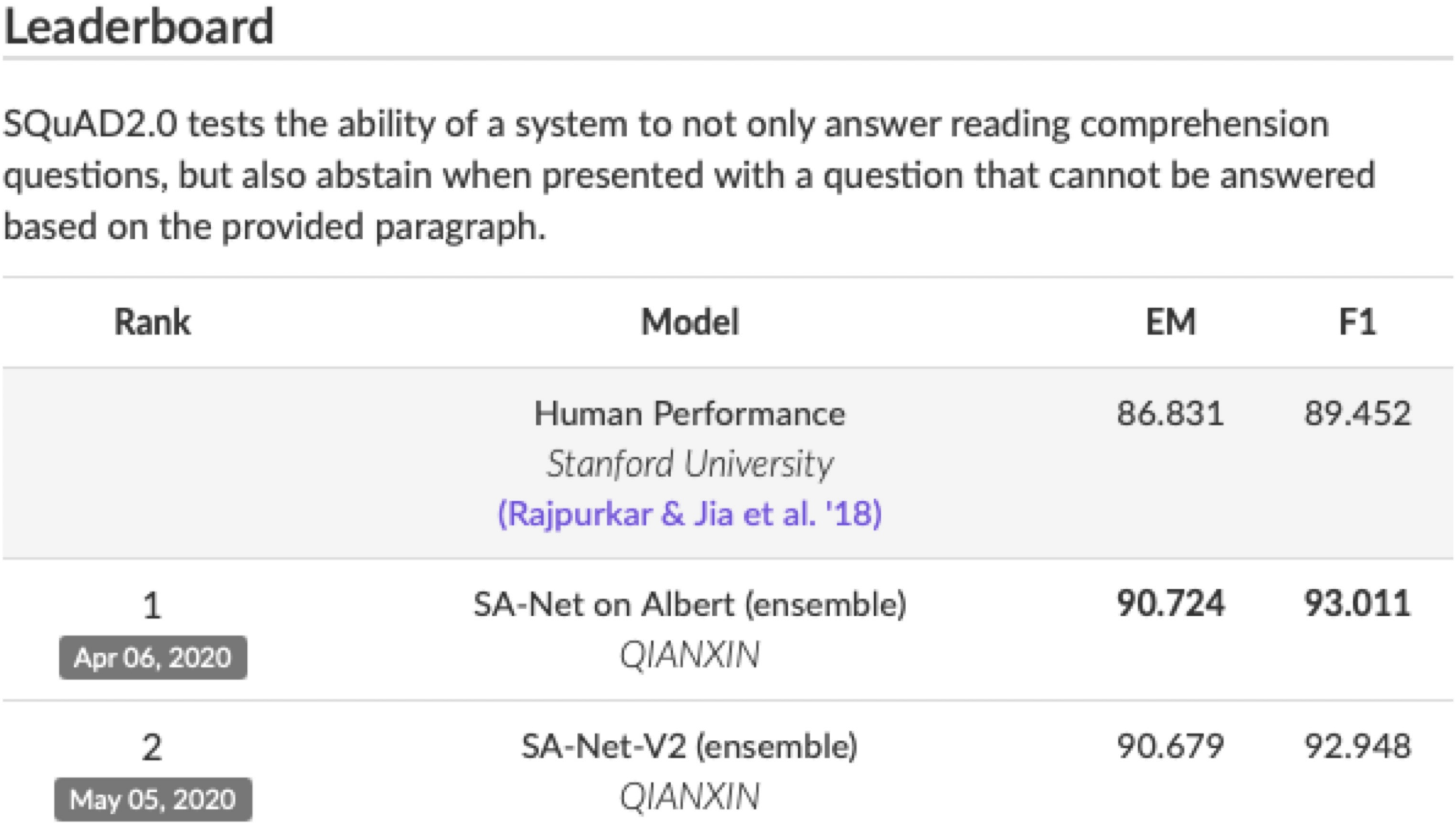
% who recover from migraines	**Drug C**	**Drug D**
Younger patients	93% (81/87)	87% (234/270)
Older patients	73% (192/263)	69% (55/80)
Both	78% (273/350)	83% (289/350)

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

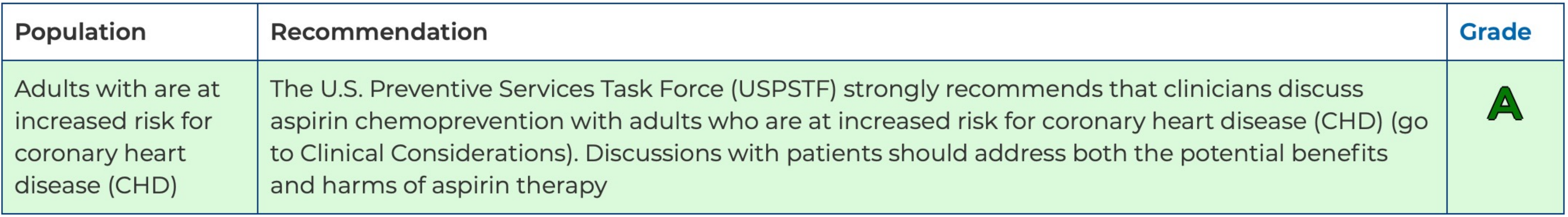
SQuAD2.0 tests the ability of a system to not only answer reading comprehension questions, but also abstain when presented with a question that cannot be answered based on the provided paragraph.

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

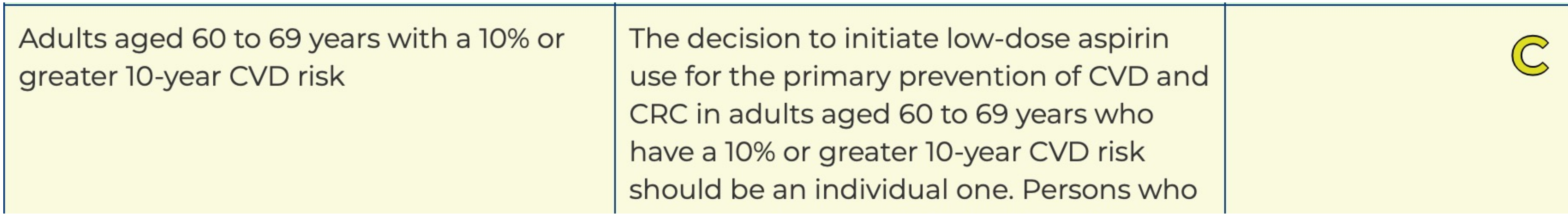
Counterfactual random variables

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

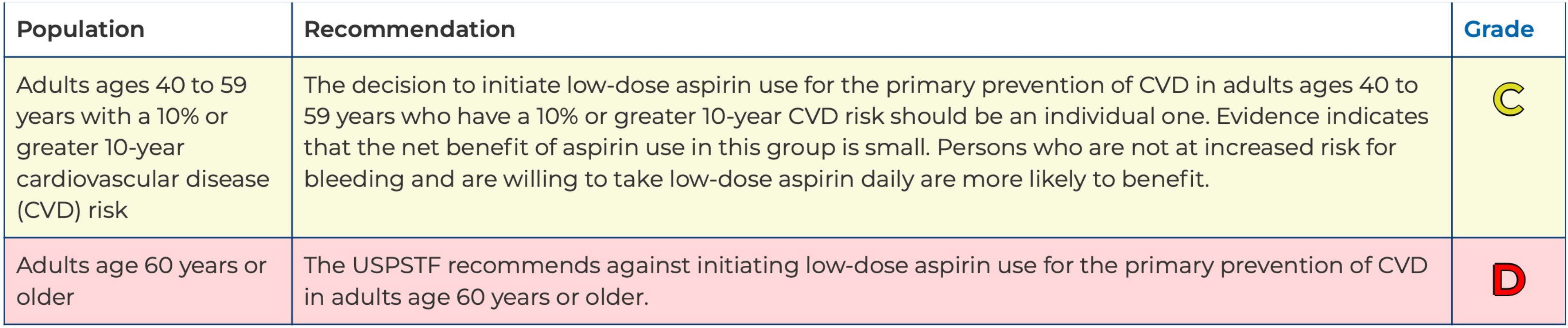
Aspirin and CVD, 2002, 2016, and 2021



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



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

Course policies

Syllabus is on Canvas, but may change

Office hours TBD: please fill out the poll (posted to Canvas)

Class will be either on Zoom or Tech M345

Course policies: Grading

Homework 50%

Split into four or five assignments

Late assignments lose 1/7th per day

Final project 40%

Split into a proposal, update, presentation, and report

Participation 10%

In-class or online discussions, anonymous surveys, peer feedback of final projects,

Tentative schedule: Weeks 1-3

1. Motivating causal inference
   - Simpson’s paradox
   - Counterfactuals
   - Randomized experiments
2. Review of fundamentals
   - Probability and statistics
   - Graphical models and conditional independence
3. Basic methods in causal inference
   - Connect potential outcomes to observational data
   - Simple confounding and identification
   - **Project proposals due**

Tentative schedule: Weeks 4-6

4. Estimators of causal effects
    - Outcome models
    - Propensity score models
5. Unmeasured confounding and identification
6. Structure learning
    - Testing for conditional independences
    - PC and GES Algorithms
    - **Project update due**

Tentative schedule: Weeks 7-10

7. Missing data
8. Measurement error and proxies
   - **Project presentations**
9. Selection bias and case-control studies
10. Evaluating causal claims in scientific literature
    - **Presentation feedback due**

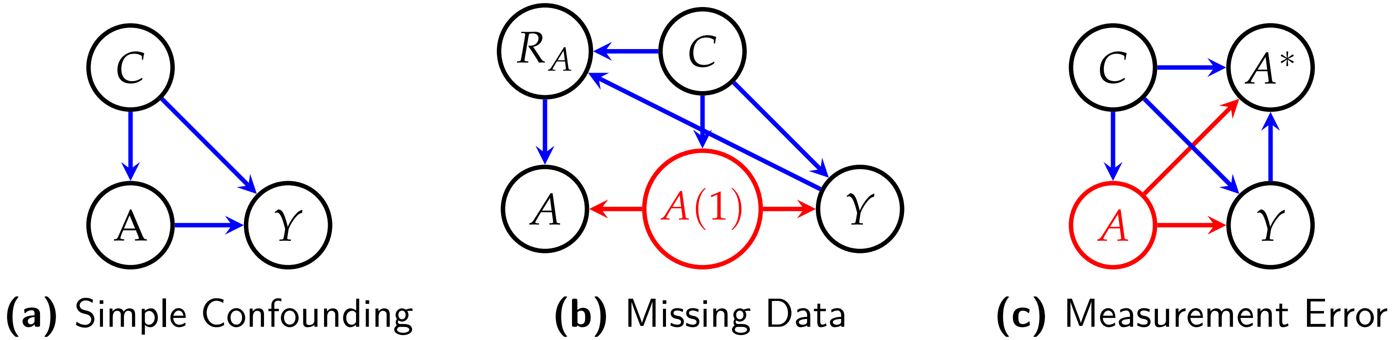
Final projects

1. Pick a dataset
   - What is the causal question you're interested in?
   - Does this dataset contain enough data to answer it?
2. Pick a graphical model that describes the data
   - Use domain knowledge and data-driven methods
3. Identification
   - Theoretical justification for why you can answer the question
4. Estimation
   - Use a method to compute the effect from data
5. Additional methods: missing data, measurement error, etc
6. Analysis of the results

For Wednesday

- Fill out the doodle for office hour availability
- Read "The C-Word: Scientific Euphemisms Do Not Improve Causal Inference From Observational Data”
- Make a GitHub account if necessary
- Start thinking about what data you might be interested in exploring for your final project

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.

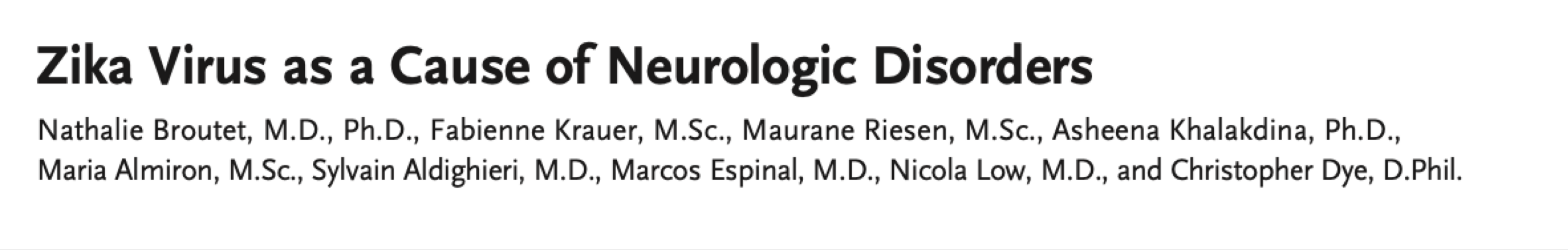
A^*	C	Y
0	1	0
0	1	1
0	0	1
1	0	1

(c) Measurement Error

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

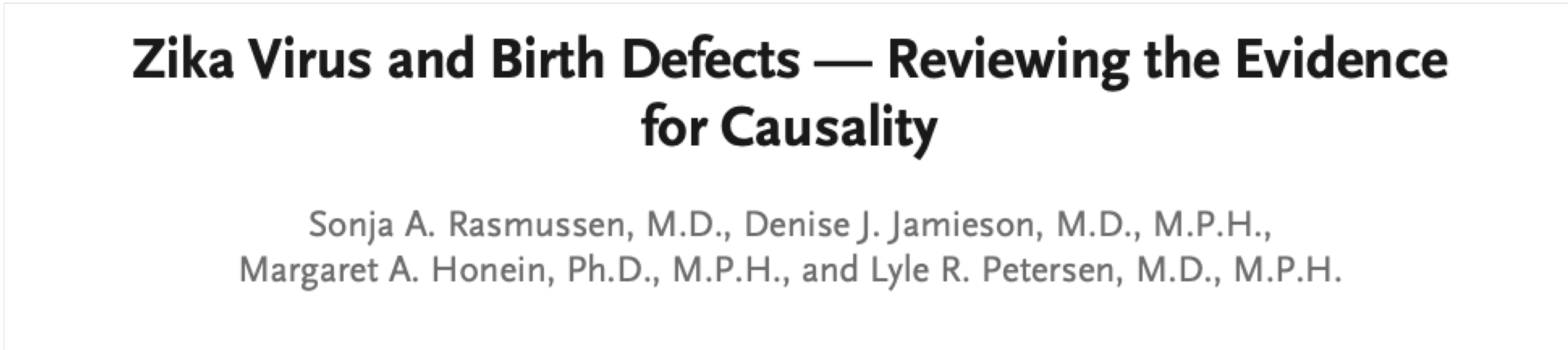
(d) Mismeasurement

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.



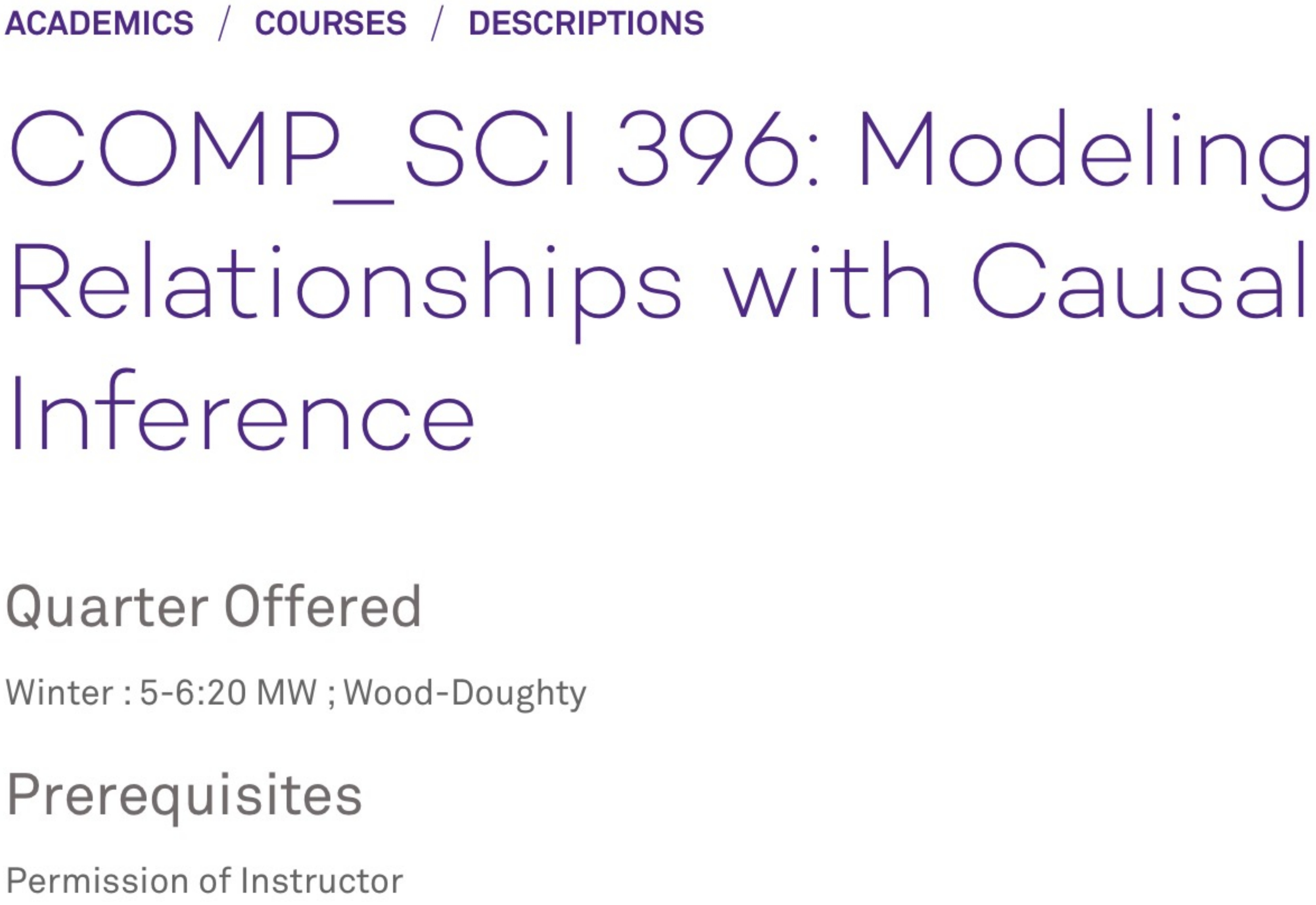
Zika Virus and Birth Defects — Reviewing the Evidence for Causality

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