Causal Inference Case Studies

Irene Y. Chen

@irenetrampoline

6.S897 / HST.956 Machine Learning for Healthcare: Recitation 6

Housekeeping

1. IBM announcement from Willie
2. Midsemester feedback form
   - We read every comment
   - Already taking suggestions into account
3. Final project mentors
   - Received initial feedback
   - Email TAs with questions
4. HW5 out
   - Mystery poll??

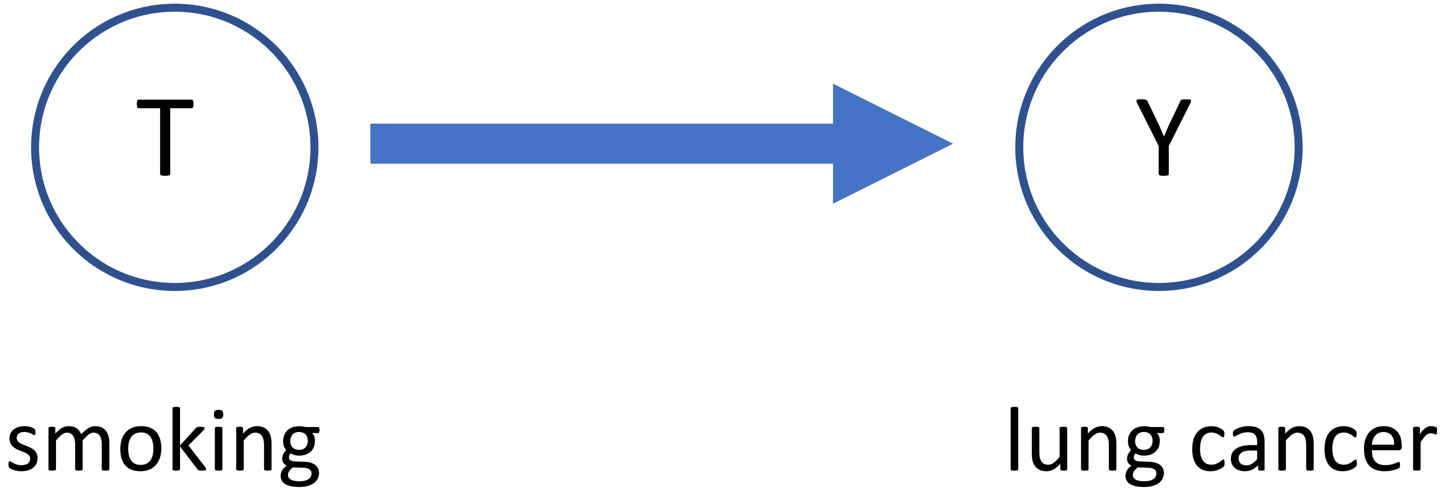
Agenda for today

1. ~~Housekeeping~~
2. Review lecture material [15 mins]
3. Post surgical opioid abuse [15 mins]
4. Diabetes treatment management [15 mins]

Goal: learn practical causality analysis tools for homework, final projects, and beyond

You inherit a tobacco company in the 1950’s

Does smoking cause cancer?

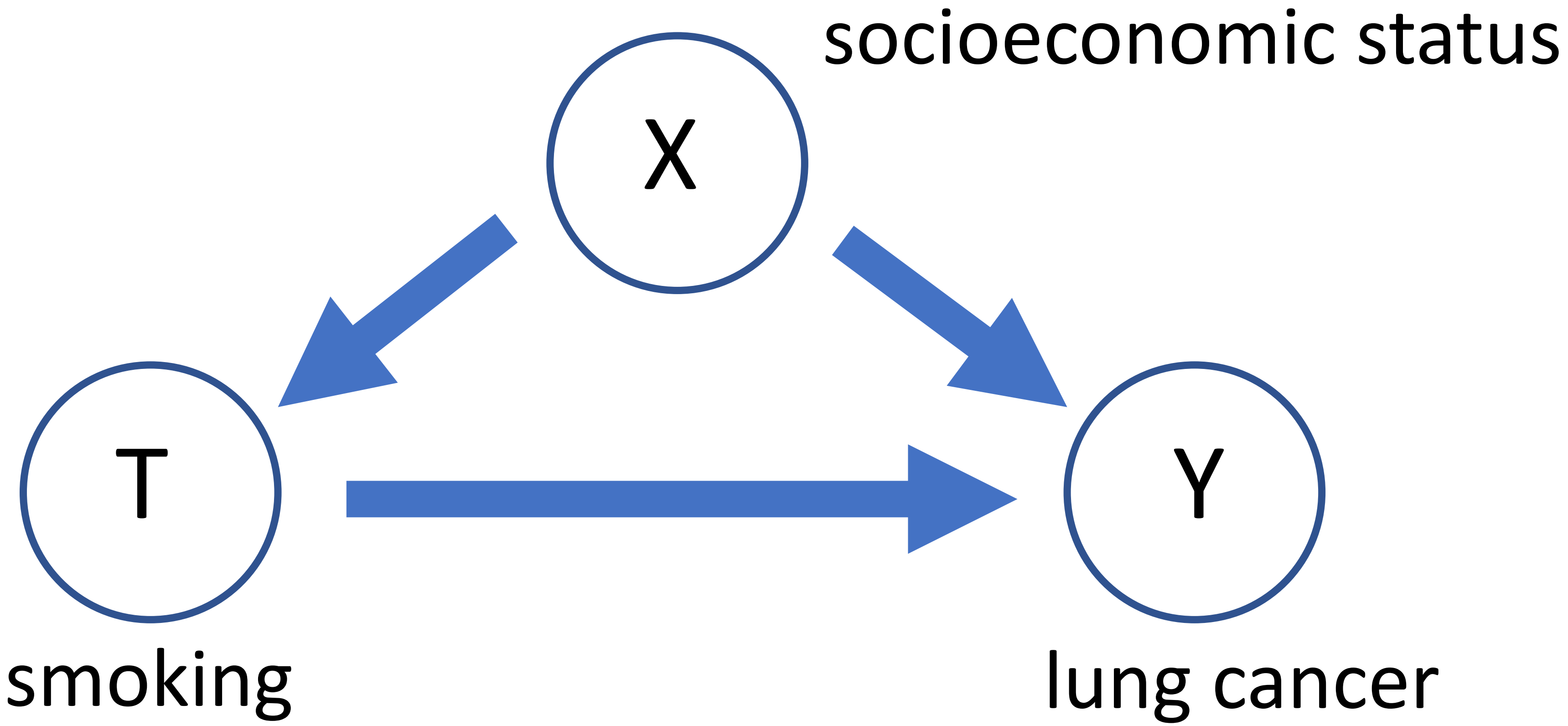


Observed association; cannot do randomized control trial

Does cancer cause smoking?

Probably not: Smoking begins years before diagnosis.

Does smoking cause cancer?

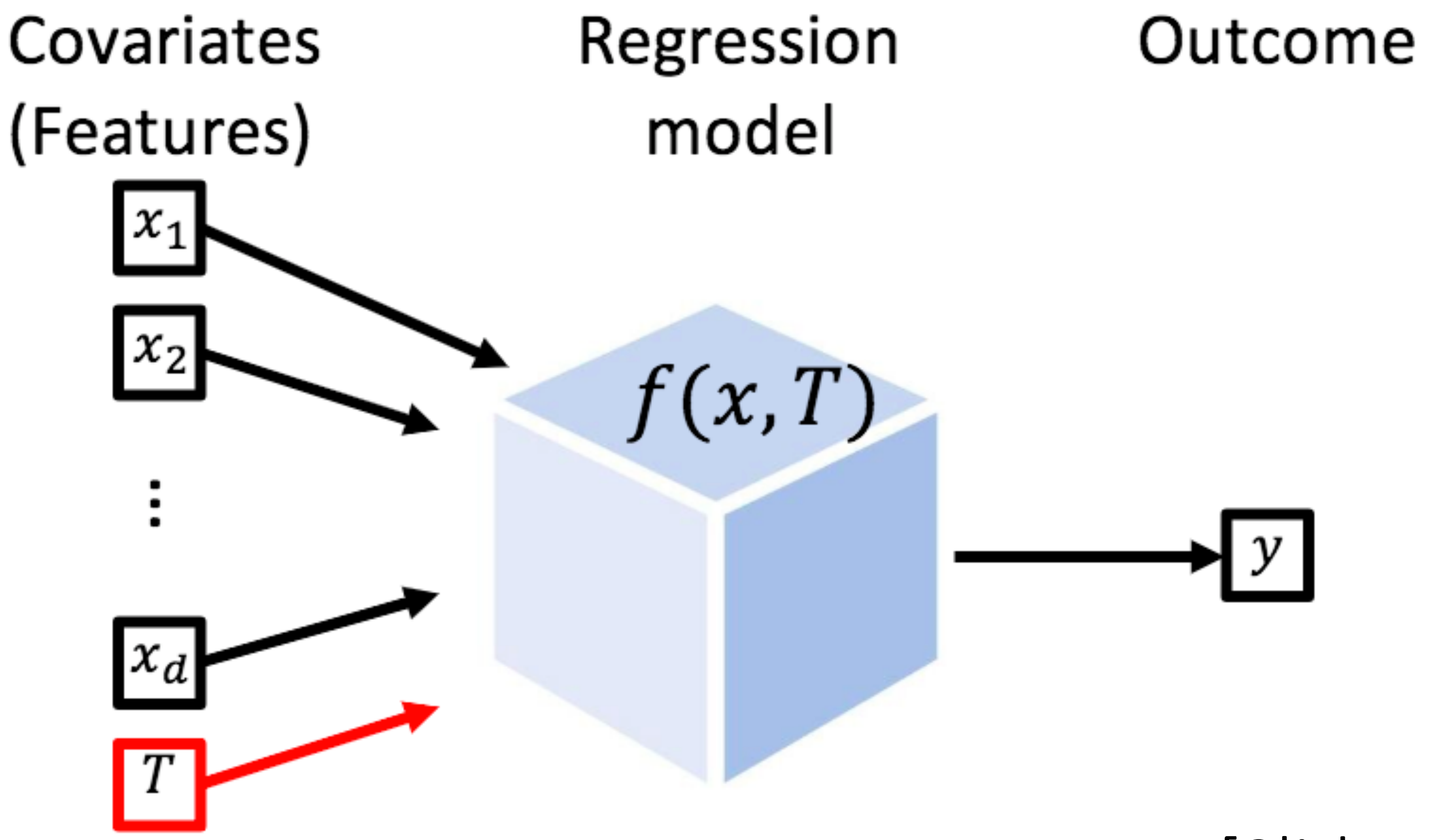


How can we estimate ATE = E[Y=1|T=1,X] - E[Y=0|T=0,X]?

Three ML methods to control for confounders

1. Covariate adjustment
2. Matching
3. Propensity scoring

Explicitly model the relationship between treatment, confounders, and outcome:



[Slide 4 of lecture 15]

Covariate adjustment (reminder)

- Under ignorability,
  $CATE(x)$ =
  $\mathbb{E}_{x\sim p(x)}\left[\, \mathbb{E}[Y_1|T=1,x] - \mathbb{E}[Y_0|T=0,x]\right]$

- Fit a model $f(x,t) \approx \mathbb{E}[Y_t|T=t,x]$, then:
  $\widehat{CATE}(x_i) = f(x_i,1) - f(x_i,0)$.

[Slide 5 of lecture 15]

Recap: Covariate adjustment

- "Plug in different values of treatment and see what happens"
- Assumes we have a very accurate and calibrated $f(x, T)$
- If data is nonlinear and we assume linearity, CATE and ATE estimates can be very misleading
- Recent research has investigated using different model classes for f, e.g. random forests and neural networks. We must then figure out how to modify the learning criteria

Matching

- Find each unit's long-lost counterfactual identical twin, check up on his outcome

Obama, had he gone to law school

Obama, had he gone to business school

[Slide 14 of lecture 15]

1-NN Matching

- Let $d(\cdot,\cdot)$ be a metric between x's
- For each i, define $j(i) = \underset{j \text{ s.t. } t_j \neq t_i}{\operatorname{argmin}} \; d(x_j, x_i)$

  $j(i)$ is the nearest counterfactual neighbor of i
- $t_i = 1$, unit i is treated:

$$\widehat{CATE}(x_i) = y_i - y_{j(i)}$$

- $t_i = 0$, unit i is control:

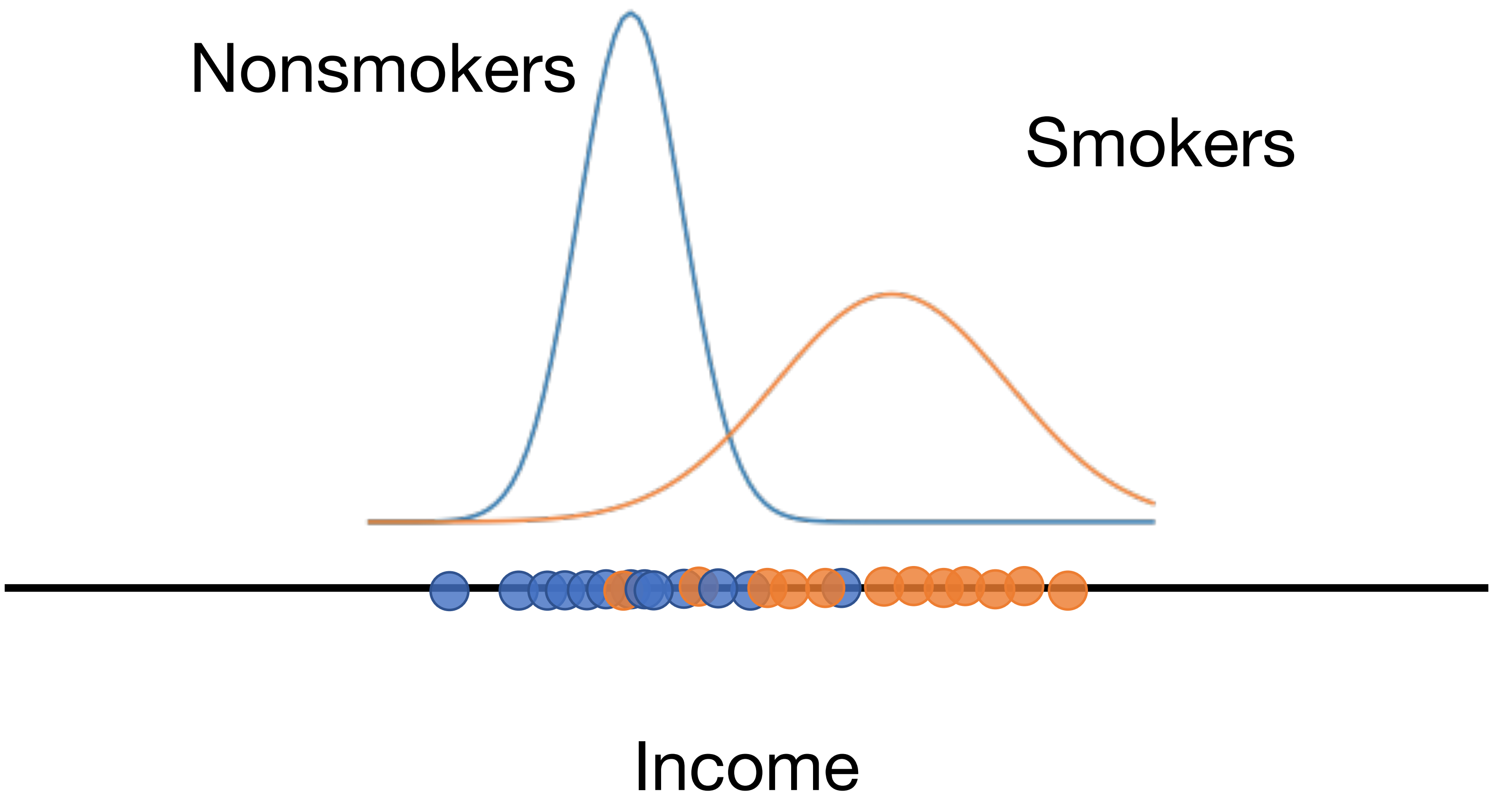
$$\widehat{CATE}(x_i) = y_{j(i)} - y_i$$

[Slide 18 of lecture 15]

Recap: Matching

- “Approximate my long lost twin and compare results”
- Useful for both CATE and ATE
- In practice, difficult because of finding right distance function and having enough twins
- Not used widely (yet!)

What if the populations are different?



Propensity score reweighting

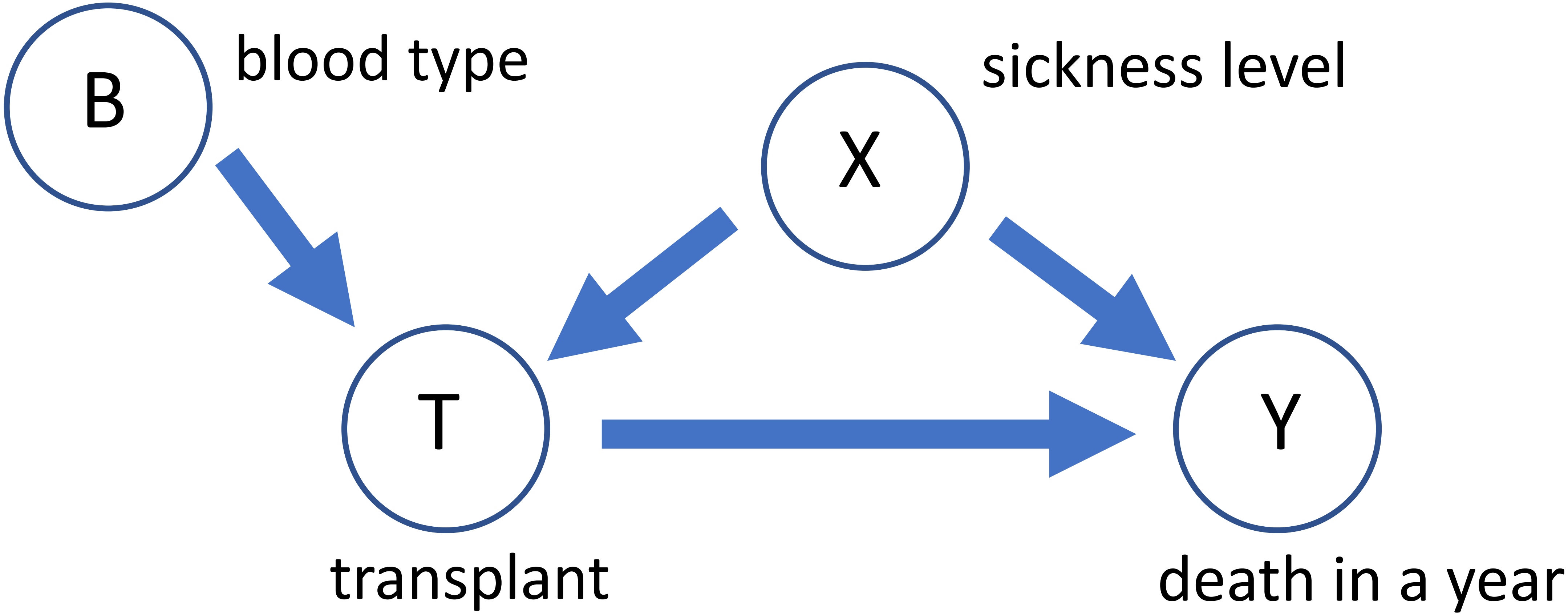
$$\hat{ATE} = \frac{1}{n} \sum_{i \text{ s.t. } t_i=1} \frac{y_i}{\hat{p}(t_i = 1|x_i)} - \frac{1}{n} \sum_{i \text{ s.t. } t_i=0} \frac{y_i}{\hat{p}(t_i = 0|x_i)}$$

[Q4 of HW5]

Recap: Propensity score reweighting

- "Give higher weight to lower represented samples"
- Can only calculate ATE
- Requires some overlap of the populations and good estimator p_hat
- If denominator is too low, weights can explode quickly

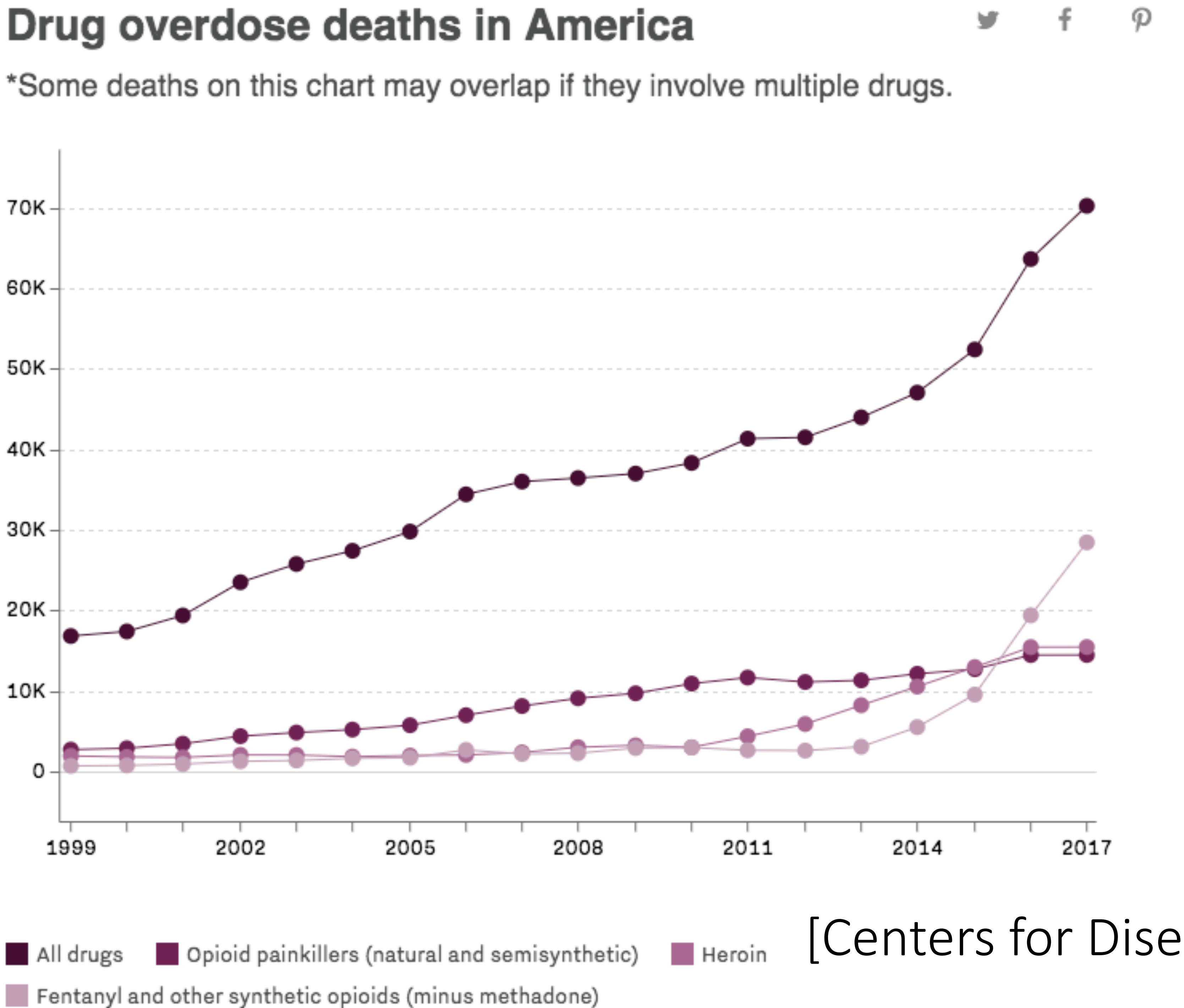
What about instrumental variables?



We can then estimate Wald estimator Cov(Y,B) / Cov(T,B)

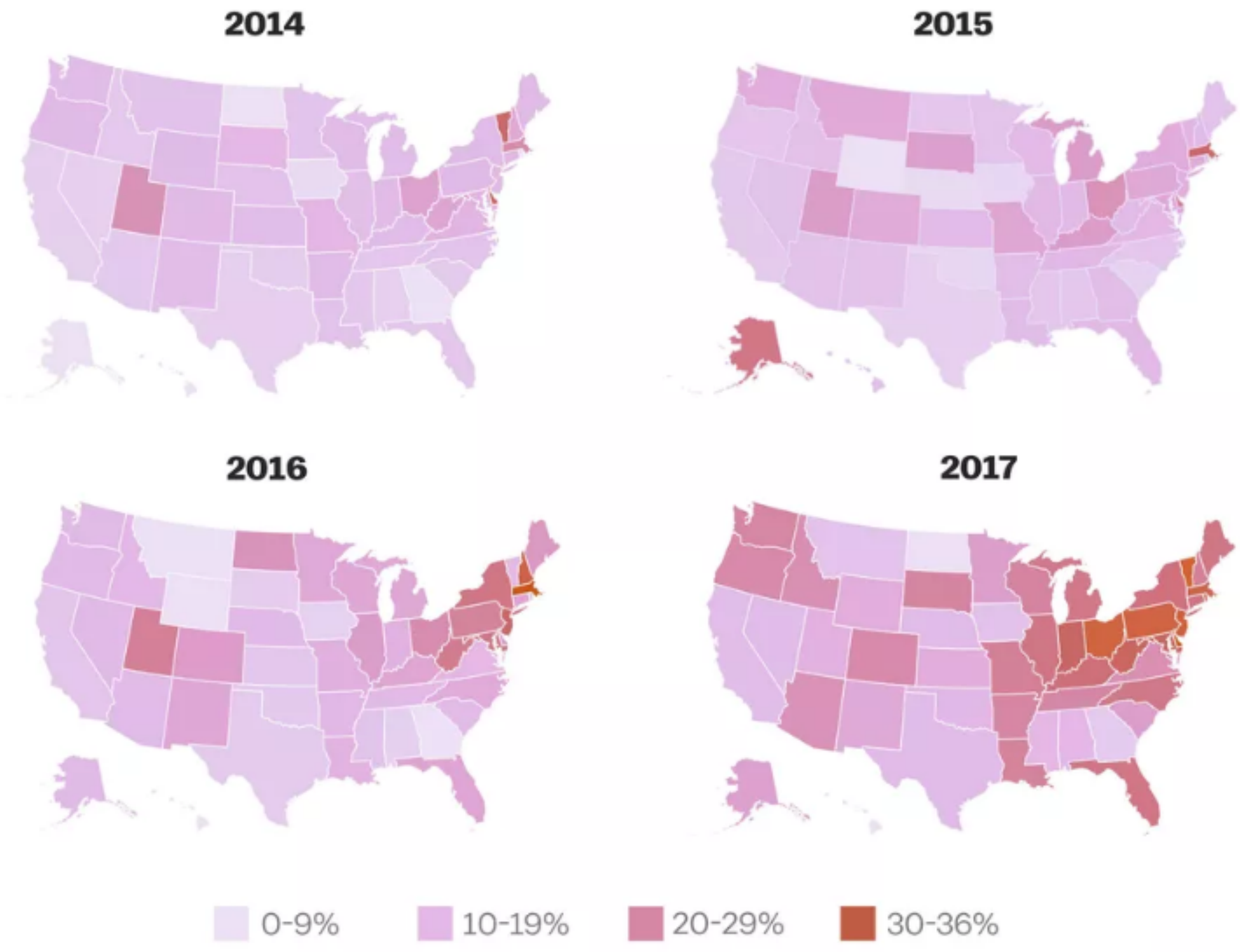
Agenda for today

1. ~~Housekeeping~~
2. ~~Review lecture material [15 mins]~~
3. Post surgical opioid abuse [15 mins]
4. Diabetes treatment management [15 mins]



[Centers for Disease Control]

Share of organ donors who died of drug overdoses



Source: Organ Procurement and Transplantation Network

Vox

Research

Postsurgical prescriptions for opioid naive patients and association with overdose and misuse: retrospective cohort study

BMJ 2018 ; 360 doi: https://doi.org/10.1136/bmj.j5790 (Published 17 January 2018)
Cite this as: *BMJ* 2018;360:j5790

Article | Related content | Metrics | Responses | Peer review

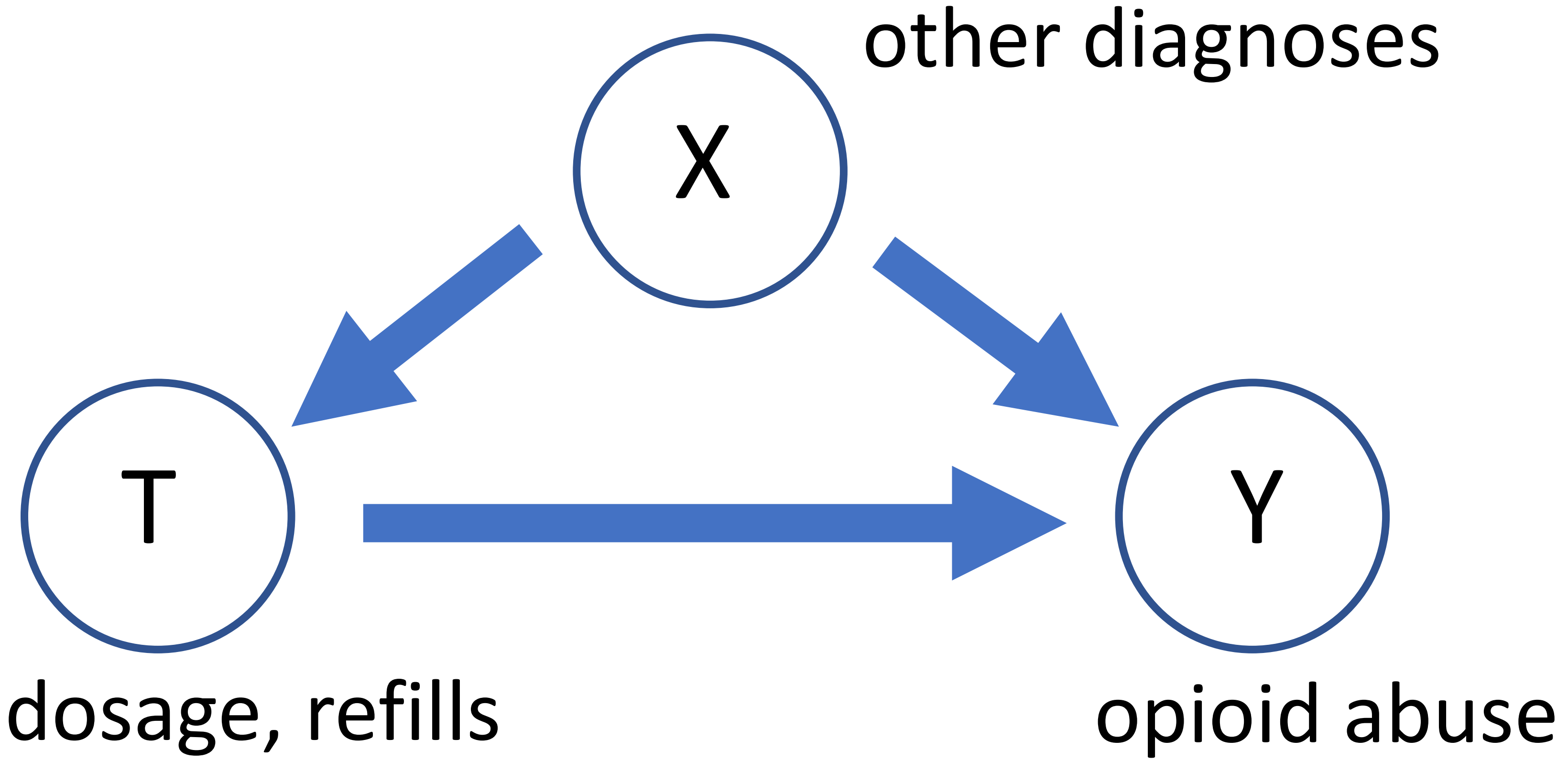
Gabriel A Brat, instructor in surgery[1 2], *Denis Agniel, postdoctoral fellow*[1], *Andrew Beam, research scientist*[1], *Brian Yorkgitis, assistant professor in surgery*[3], *Mark Bicket, assistant professor in anesthesia*[4], *Mark Homer, postdoctoral fellow*[1], *Kathe P Fox, director*[5], *Daniel B Knecht, chief of staff*[5], *Cheryl N McMahill-Walraven, director*[5], *Nathan Palmer, research scientist*[1], *Isaac Kohane, department chair*[1]

Author affiliations

Correspondence to: G A Brat gbrat@bidmc.harvard.edu

Accepted 1 December 2017

Do postsurgical opioids cause opioid abuse?



Aetna Insurance claims

Pros

- Complete patient record
- Hospital and pharmacy care
- Surgical claims from CPT, outcomes from ICD-9 codes

Cons

- Lacking granular information about hospital stays (e.g. lab values)
- CPT and ICD-9 codes can be incorrect or manipulated for billing purposes

Data source

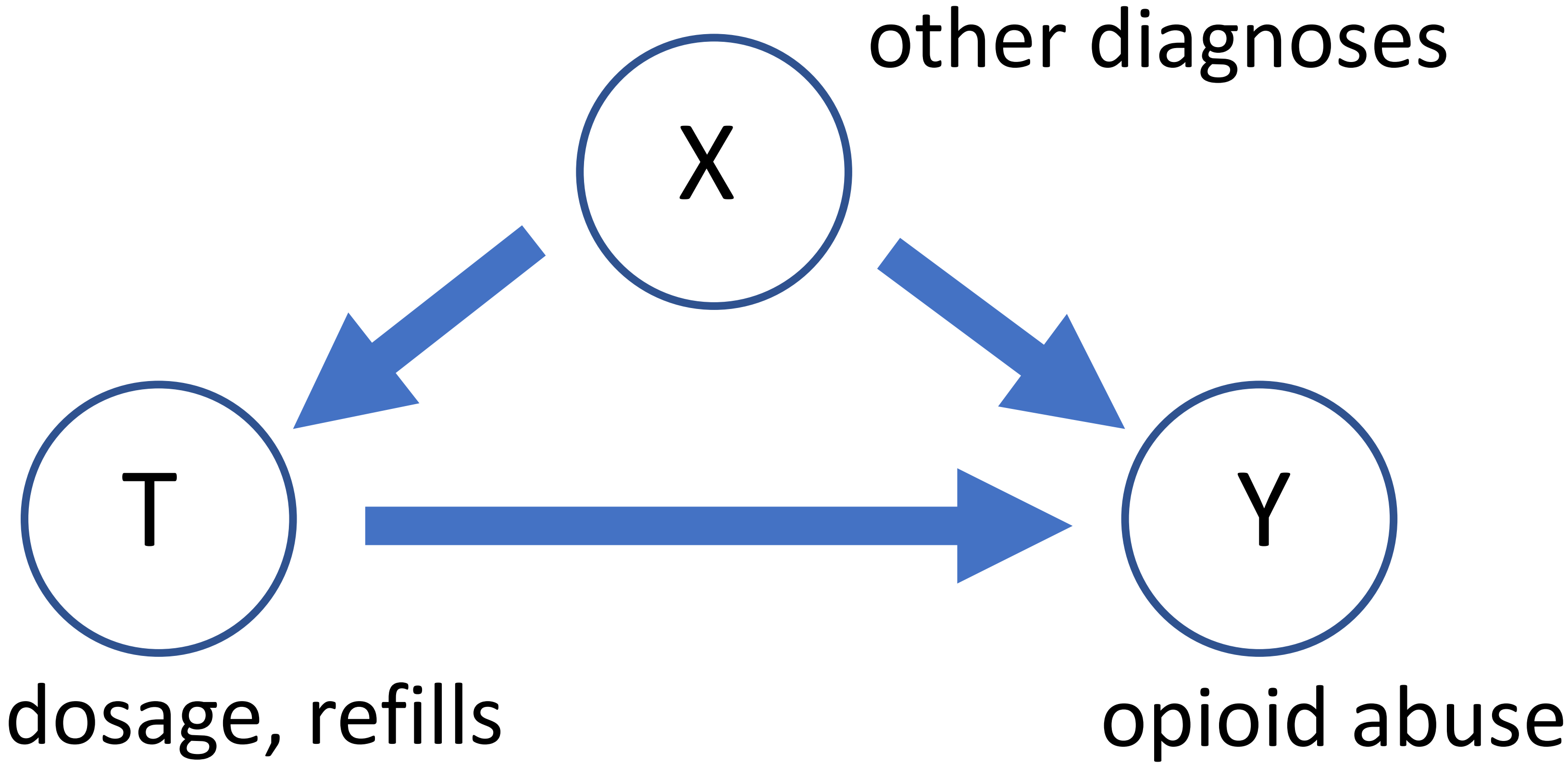
Include

- Patients with “complete” medical and pharmacy insurance records
- Underwent first surgery
- Opioid naïve: little/no previous opioid use

Final cohort

- Large dataset (37 million)
- Longitudinal (2008-2016)
- After inclusion criteria, 1 million opioid naïve patients undergoing surgery

Do postsurgical opioids cause opioid abuse?



How do we define T, Y, and X?

What is treatment T?

- Refill
- Total dosage
- Duration of use

What is outcome Y?

- ICD-9 code for opioid dependence, abuse, and overdose
- Only include diagnosis codes related to prescription opioids

What are confounders X?

- Demographics (age, sex)
- US state of residence
- surgery type group
- surgery year
- presurgical diagnoses

Statistical analysis

- Weighted linear regression for log transformed weekly rates of misuse
  - Each week weighted according to sample size
  - Create outcome of adjusted analysis of **time until misuse event** using Cox proportional hazards (survival analysis!)
  - Results report **multiplicative percentage increases** in rate
- Sensitivity analysis to rule out structural confounders
  - Interaction term between duration and year indicator
  - Interaction between duration and state of residence indicator
  - Build in an unobserved confounder with a Bernoulli random variable

Recap: Postsurgical opioid use to misuse

- “Duration more than dosage use may cause opioid misuse”
- Use covariate adjustment to estimate multiplicative effects
- Interaction terms

Agenda for today

1. ~~Housekeeping~~
2. ~~Review lecture material [15 mins]~~
3. ~~Post surgical opioid abuse [15 mins]~~
4. Diabetes treatment management [15 mins]

Clinical Care/Education/Nutrition/Psychosocial Research

Personalized Diabetes Management Using Electronic Medical Records

Dimitris Bertsimas, Nathan Kallus, Alexander M. Weinstein and Ying Daisy Zhuo

+ Author Affiliations

Corresponding author: Dimitris Bertsimas, dbertsim@mit.edu.

Diabetes Care 2017 Feb; 40(2): 210-217.
https://doi.org/10.2337/dc16-0826

Check for updates

Previous Next

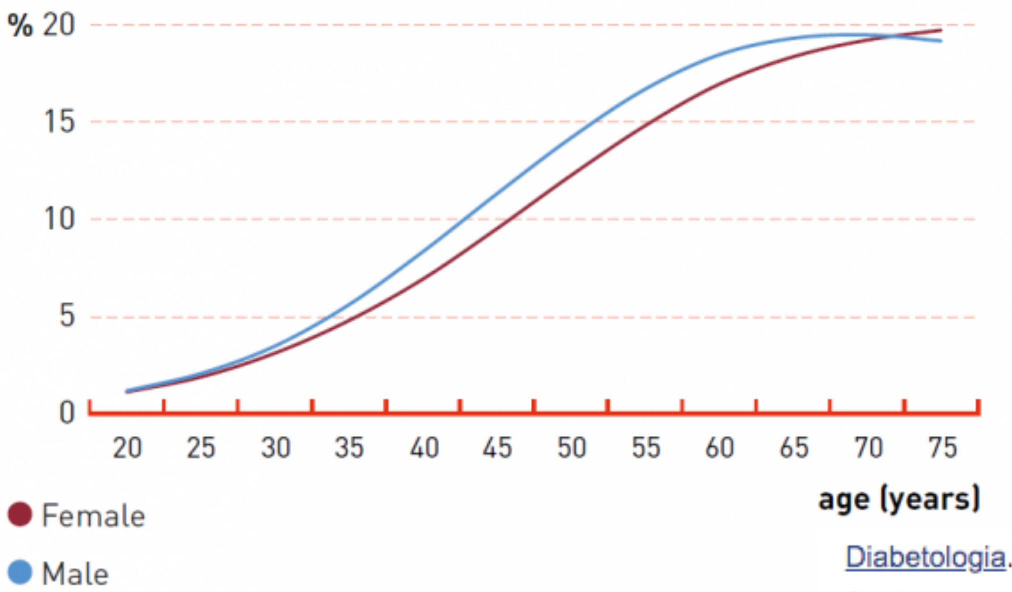
Article | Figures & Tables | Suppl Material | Info & Metrics | PDF

Abstract

OBJECTIVE Current clinical guidelines for managing type 2 diabetes do not differentiate based on patient-specific factors. We present a data-driven algorithm for personalized diabetes management that improves health outcomes relative to the standard of care.

Type 2 Diabetes Treatment Still a Mystery

Figure 2.2 Prevalence (%) of people with diabetes by age and sex, 2013



BMJ Open. 2015; 5(5): e007375.
Published online 2015 May 12. doi: 10.1136/bmjopen-2014-007375
PMCID: PMC4431069
PMID: 25967997

Racial ethnic differences in type 2 diabetes treatment patterns and glycaemic control in the Boston Area Community Health Survey

Sunali D Goonesekera, May H Yang, Susan A Hall, Shona C Fang, Rebecca S Piccolo, and John B McKinlay

‣ Author information ‣ Article notes ‣ Copyright and License information Disclaimer

Diabetologia. Author manuscript; available in PMC 2014 Dec 1.
Published in final edited form as:
Diabetologia. 2013 Dec; 56(12): 10.1007/s00125-013-3078-7.
Published online 2013 Oct 5. doi: 10.1007/s00125-013-3078-7
PMCID: PMC3842214
NIHMSID: NIHMS529351
PMID: 24092493

Age-related differences in glycaemic control in diabetes

Elizabeth Selvin[1] and Christina M. Parrinello[1]

What do we include in this analysis?

Inclusion criteria

- Patients in hospital EMR for >1 year
- Prescription for at least one blood glucose regulation agent
- At least three recorded laboratory results for HbA1C
- No recorded diagnosis of type 1 diabetes (from ICD-9 code 250.x1 or 250.x3)

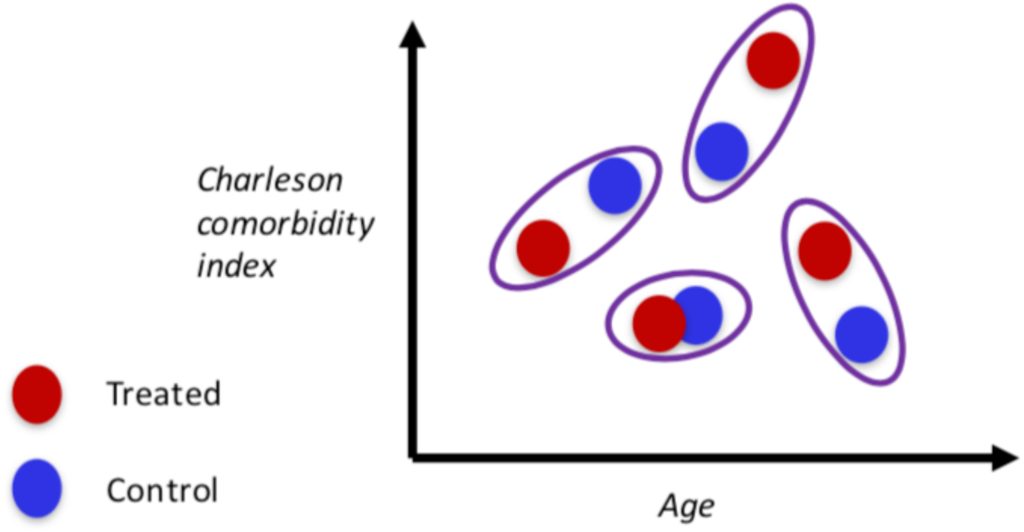
Final cohort

- 10k patients, 48k patient visits
- Access to demographic information
- Analyze all associated EMR data

What makes two patients similar or different?

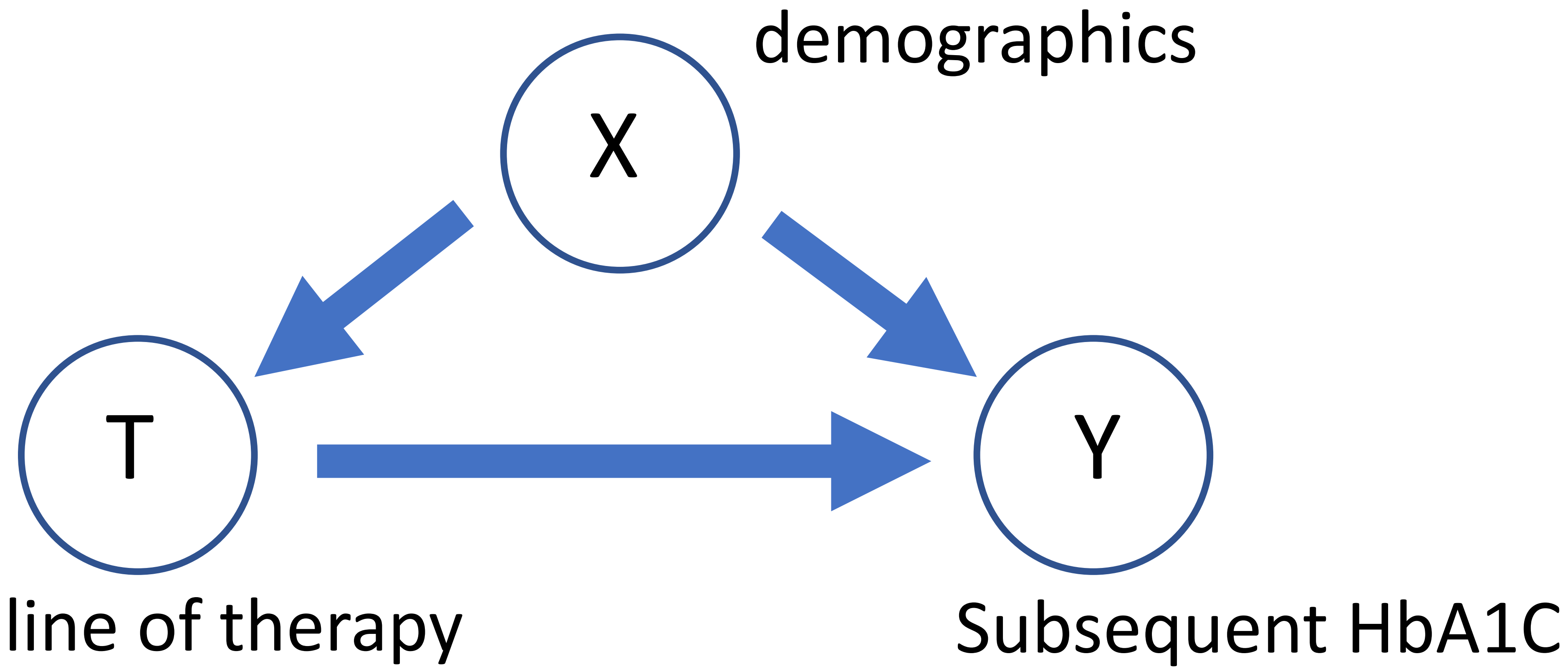
Features

- Differentiate 13 lines of therapy
- Patient visit every 100 day and average HbA1C after visit (75-200 days after)
- Collect what standard of care was actually administered



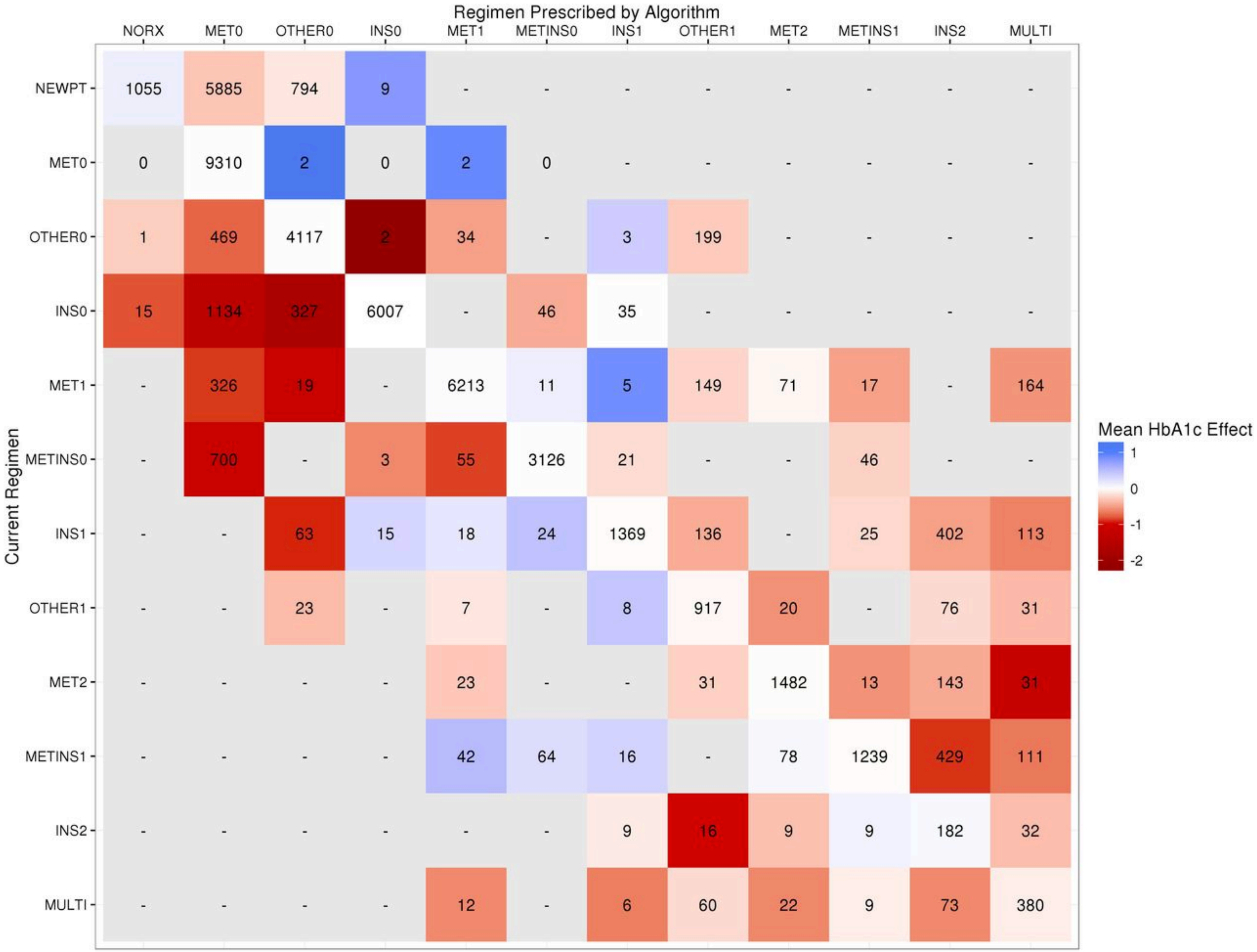
[Slide 17 of lecture 15]

Which treatment will lead to lower HbA1C?

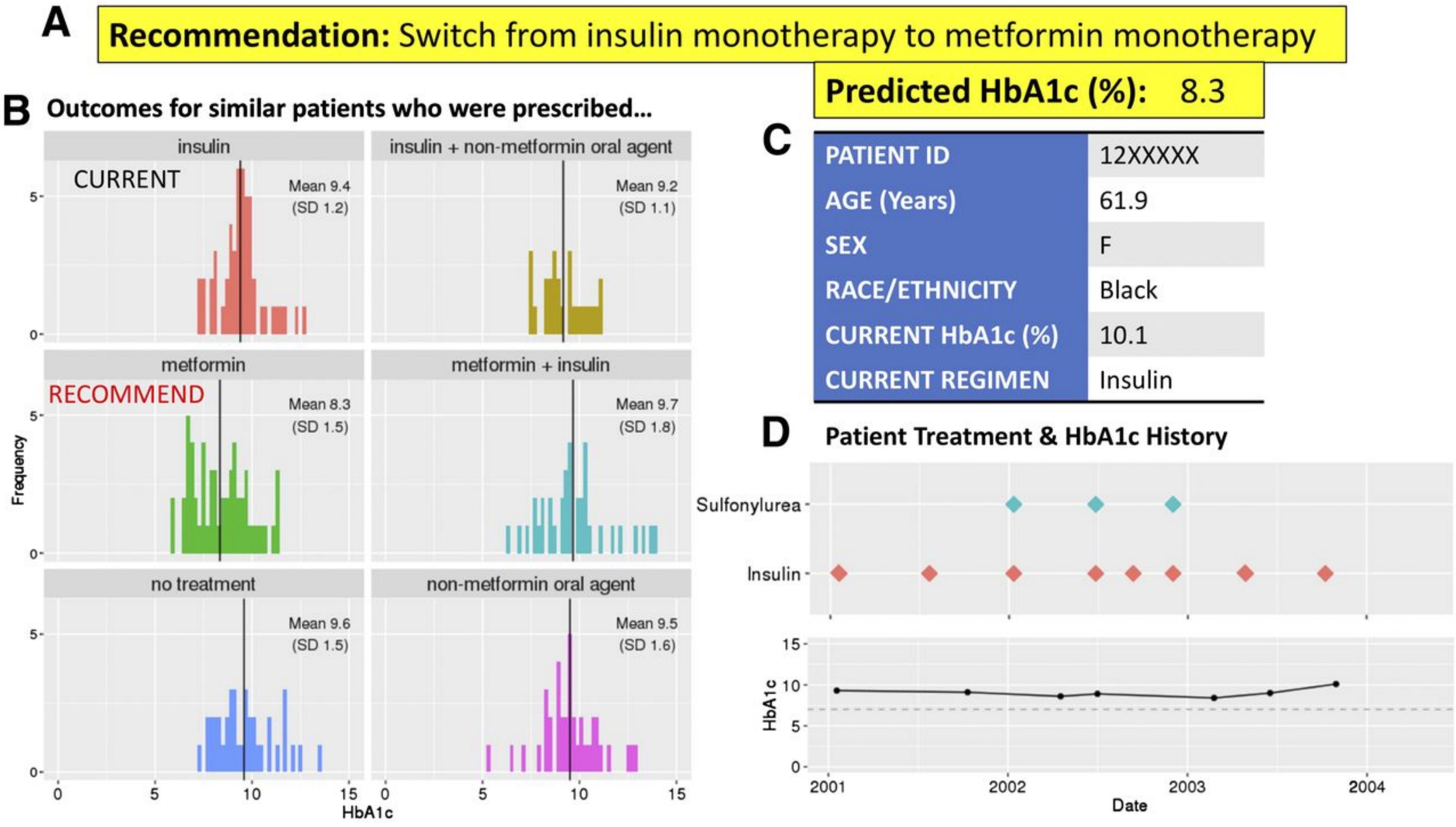


Model

- For each patient visit, find kNN regression to predict HbA1C under every possible treatment
- Algorithm prescribes regimen with best predicted outcome if predictive improvement exceeds threshold
- Evaluation compared actual treatment and outcome with recommended therapy and outcome
- Sensitivity analysis by drawing new training and testing splits



[Figure 1 of Bertsimas et al, 2017]



[Figure 2 of Bertsimas et al, 2017]

Recap: Diabetes treatment management

- “kNN over patients can recommend diabetes treatments”
- Use matching to estimate different treatment effects
- Evaluate by comparing predicted and actual treatment and HbA1C values
- Sensitivity analysis through repeated sampling of training and test data

Have a great weekend!