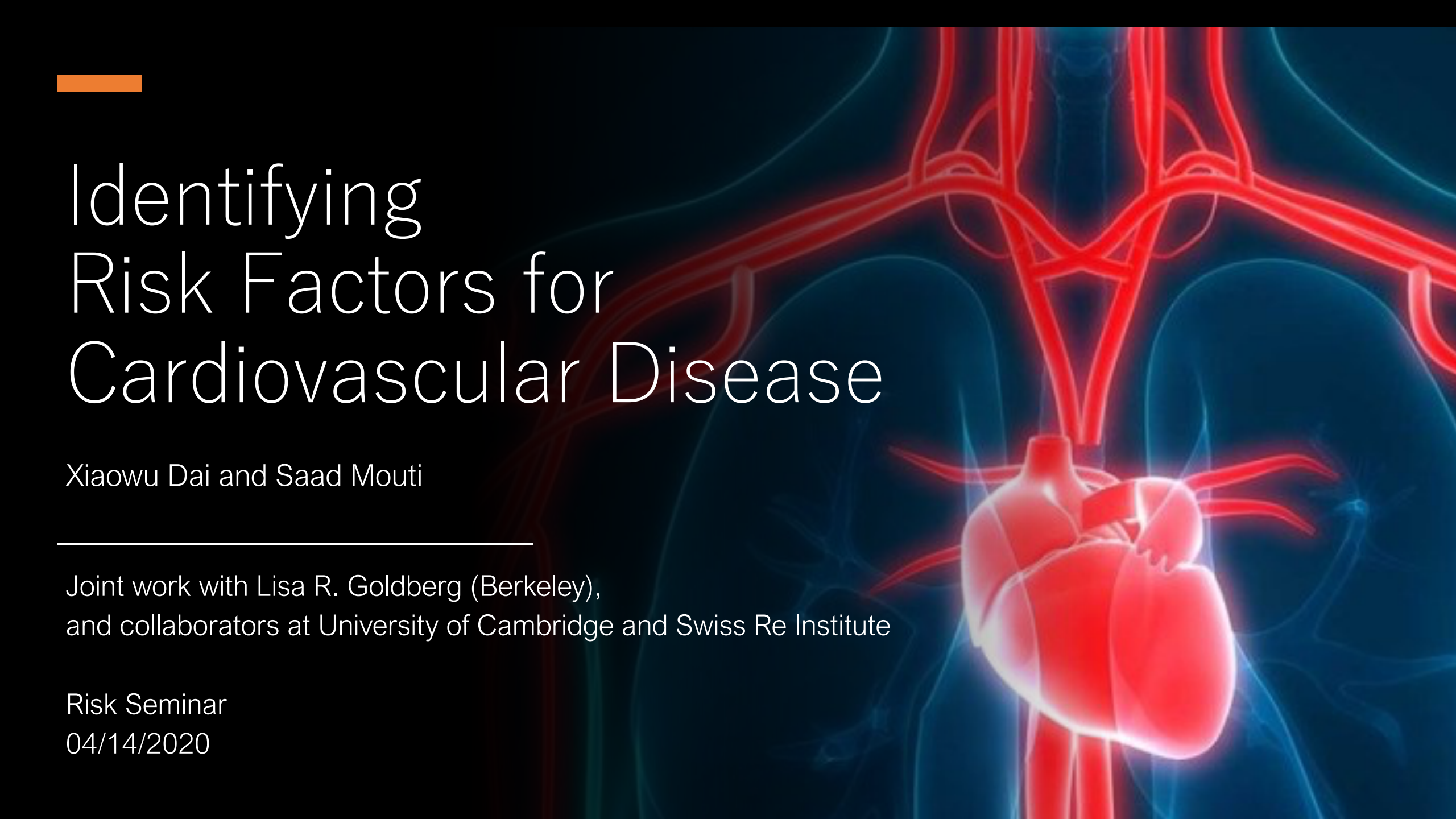
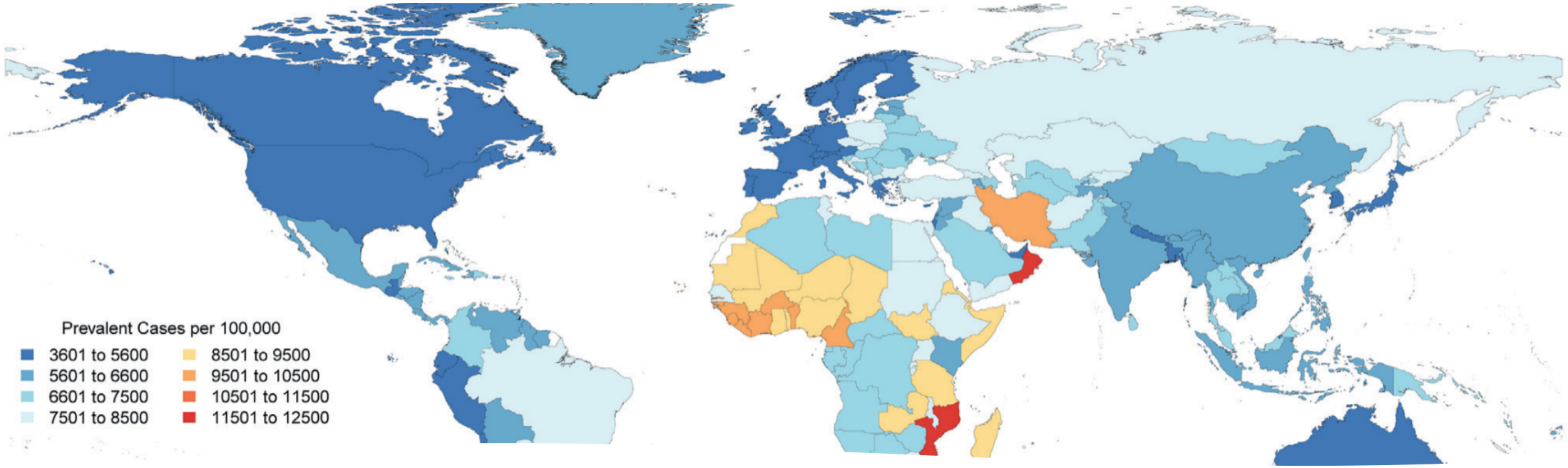
Identifying Risk Factors for Cardiovascular Disease

Xiaowu Dai and Saad Mouti

Joint work with Lisa R. Goldberg (Berkeley),
and collaborators at University of Cambridge and Swiss Re Institute

Risk Seminar
04/14/2020





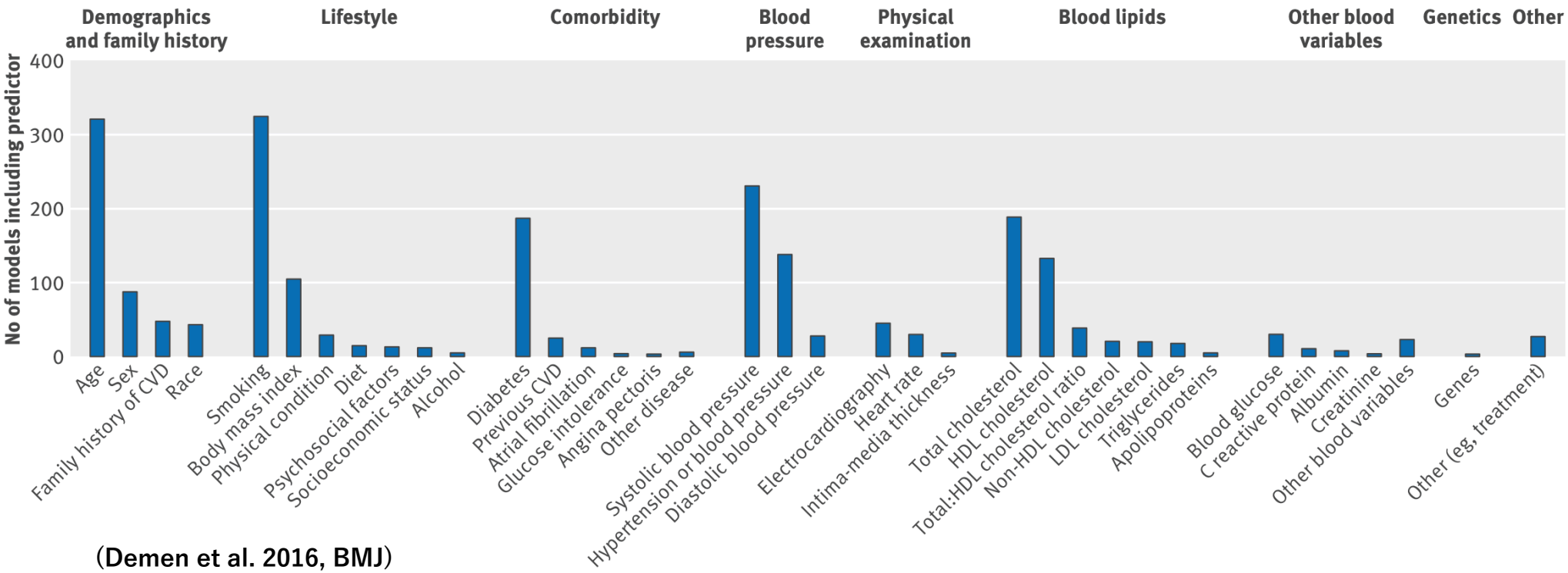
(Roth et al. 2017, JOURNAL OF THE AMERICAN COLLEGE OF CARDIOLOGY)

Cardiovascular Disease (CVD)

- A leading cause of death in the world
- United Nations recognized CVD as a major concern for global health
- Identification of CVD related risk factors is a health priority
- Individual level interventions (WHO 2017)

Goal: Statistical Methods for Identifying Risk Factors for CVD

- **Demographics** (age, gender, race, etc.)
- **Lifestyle** (smoking, bmi, diet, alcohol, etc.)
- **Comorbidity** (diabetes, other disease, etc.)
- **Blood pressure, blood lipids, genetics, stress, etc.**

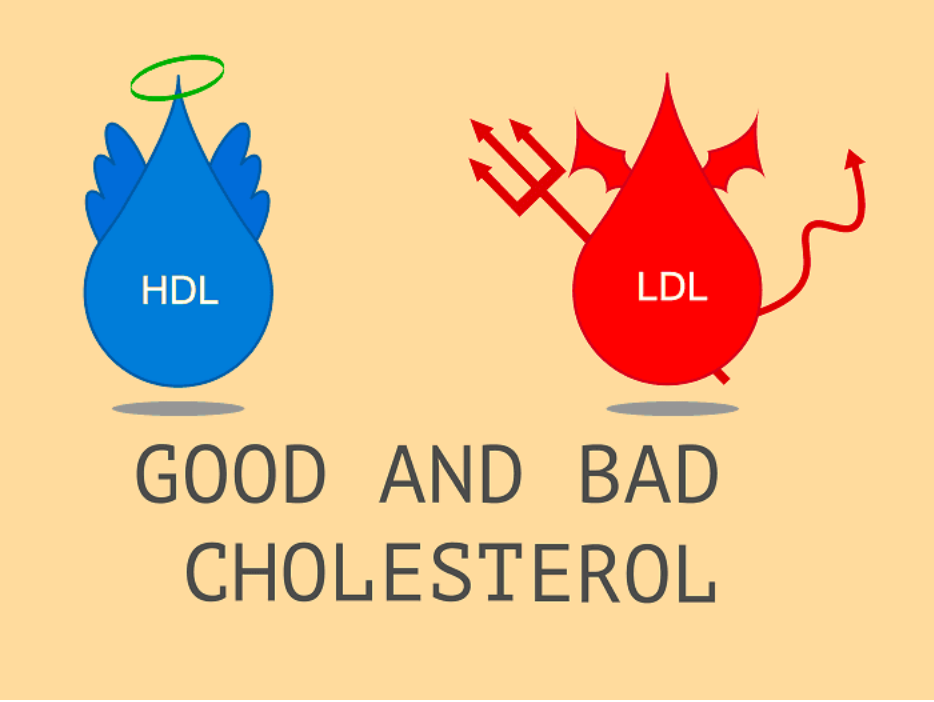


(Demen et al. 2016, BMJ)

Spurious Correlation

- High-density lipoprotein (HDL) cholesterol of higher levels is associated with low CVD risk
- But drugs that increase HDL (e.g., trapibs) do not significantly reduced CVD
- HDL is an indirect / surrogate marker, and HDL does not participate in causing or alleviating CVD
(Voight et al. 2012, The Lancet)

Our Approach: Causal inference + Subsampling



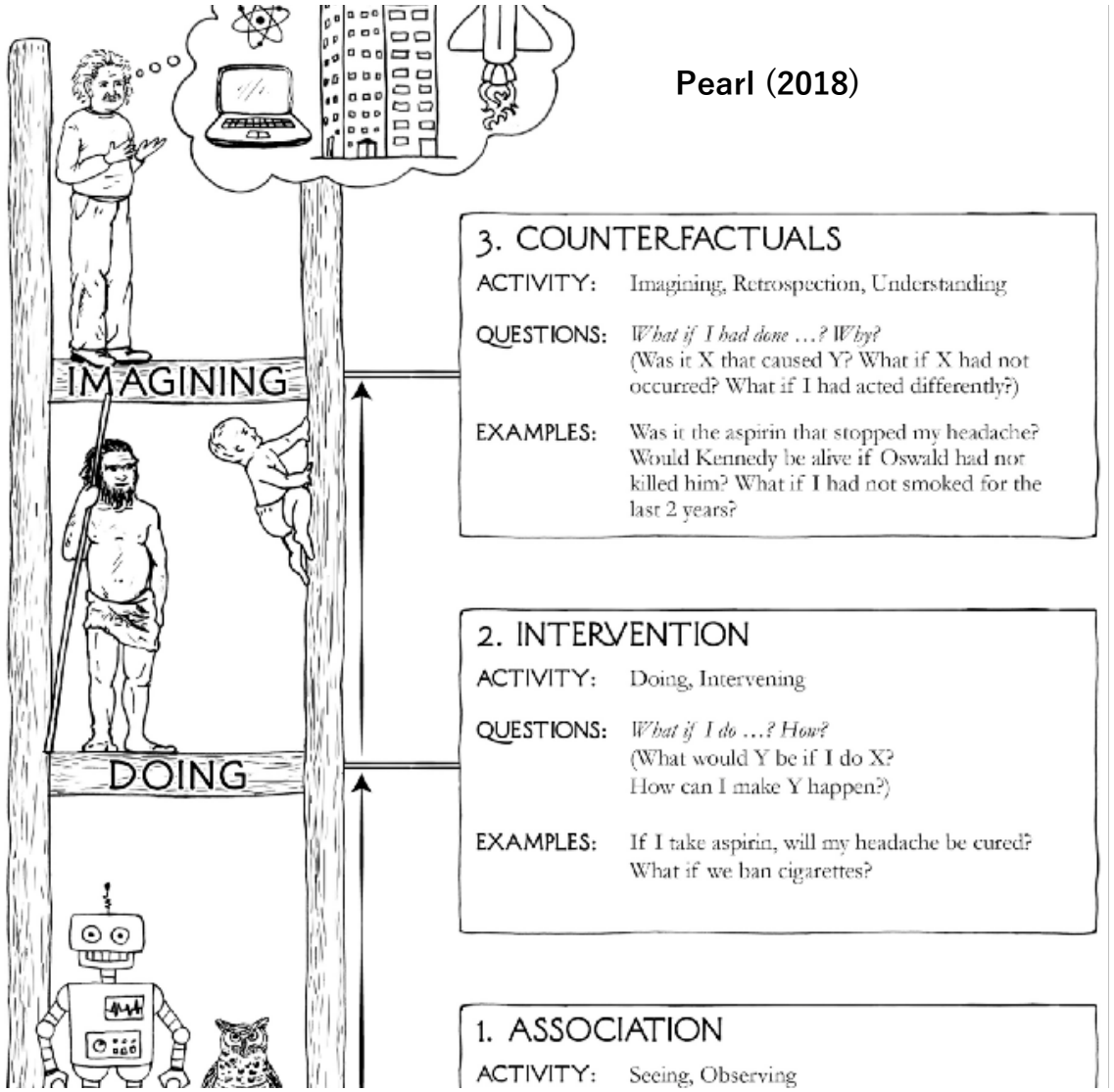
Causal Inference for CVD

Causal inference → **Personalized** therapies in cardiology

Related Work: Mendelian Randomization
(van der Kann et al., 2016, J AM Coll Cardiol;
Shameer et al., 2018, Heart)

- **Genetic** material is randomly inherited
- Alcohol intake (**Holmes et al. 2014, BMJ**)
- LDL cholesterol (**Holmes et al. 2015, Eur Heart J**)

Our Focus: Demographics, Lifestyle, Comorbidity, Blood pressure, Blood lipids.



Pearl (2018)

Q1: Does obesity cause type 2 diabetes?

Matching Method

--A method for causal inference using observational data

- **Potential outcome:** $\mathbb{E}(Y(1)|X) - \mathbb{E}(Y(0)|X)$

  $Y(1)$: outcome if receiving the treatment
  $Y(0)$: outcome if receiving the control

- **Average treatment effect** (ATE): effect for all individuals including both treatment and control

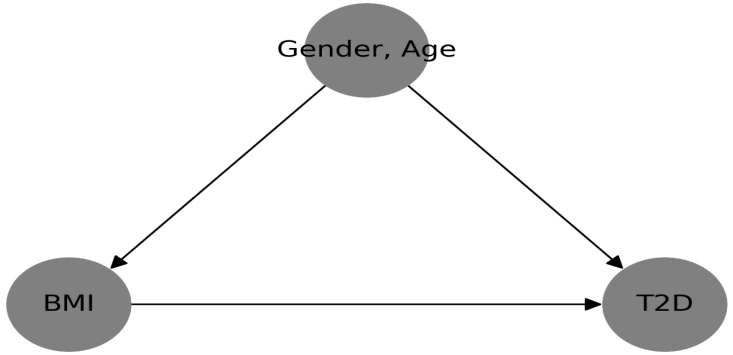
- **Assumption 1**: **Stable unit treatment value assumption** (Rubin, 1980)

  Outcomes of one individual are not affected by treatment assignment of any other individuals

- **Assumption 2: <u>Strong ignorable treatment assignment condition</u>** (Rosenbaum and Rubin, 1983)

  (2.1) *Treatment assignment is independent of the potential outcomes given the covariates*

  (2.2) *Non-vanishing probability of receiving each treatment for all values of covariates*



<u>Goal of matching method</u>: replicate a randomized experiment as closely as possible by obtaining treated and control groups with similar covariate distributions

Dr. David Unwin, MD

Dr. David Unwin, MD, is an award-winning general practitioner (or family doctor) known for pioneering the low-carb approach in his profession in the UK.

In 2016 he won the prestigious NHS Innovator of the Year award for his work with diabetes patients. On top of that, Dr. Unwin is the medical advisor at the popular Low Carb Program and is doing his best to spread knowledge about low carb among doctors, dietitians and nurses.

In 2017/18, his practice saved £57,000 on drugs for type 2 diabetes, hypertension and other conditions by offering patients a dietary alternative to medications.

Data Collection

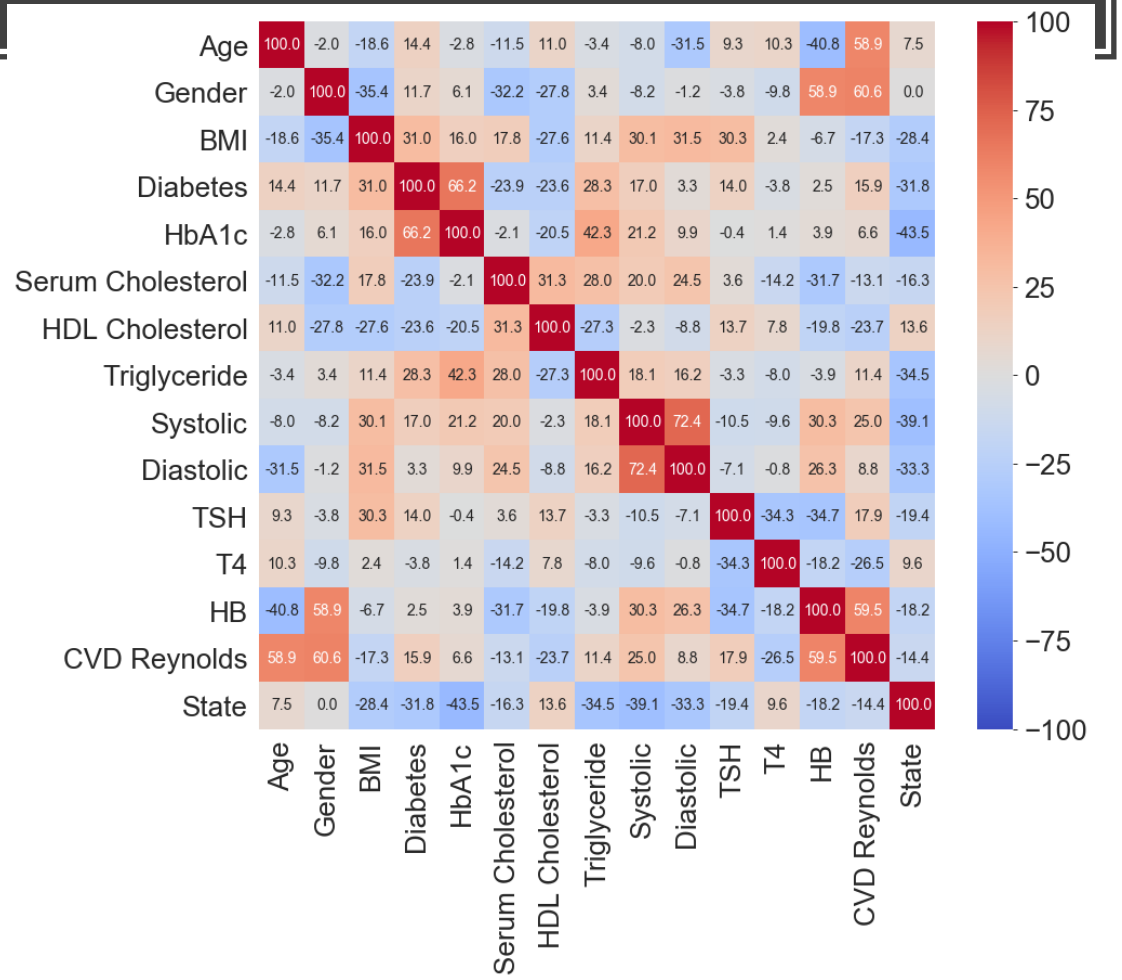
- 256 subjects in London, UK
- Dr. Unwin's office collect data for two time points
- Time 0: without low-carb diet
- Time 1: with low-card diet

Data Summary

Summary Statistics

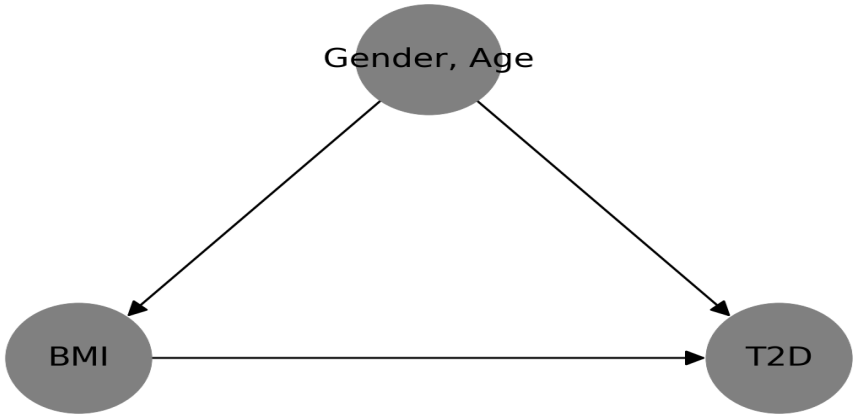
State	Variable	count	mean	std	min	25%	50%	75%	max
0	Gender	256	0.590	0.493	0.000	0.000	1.000	1.000	1.000
	Age	256	61.574	12.111	23.000	53.000	60.000	71.000	91.000
	Height	75	1.706	0.092	1.473	1.625	1.720	1.770	1.900
	Weight	251	96.160	18.621	55.300	83.700	95.000	107.000	159.000
	BMI	66	33.887	6.071	21.660	29.890	33.495	36.980	57.100
	Diabetes	256	1.281	0.811	0.000	1.000	2.000	2.000	2.000
	HbA1c/ mmol/mol	202	61.376	20.652	37.000	45.000	54.500	71.000	135.000
	Gamma-G.T Level/ U/L	137	78.664	78.426	14.000	35.000	51.000	95.000	489.000
	Serum Cholesterol	176	5.314	1.302	2.500	4.385	5.200	6.225	9.300
	HDL Cholesterol	195	1.280	0.421	0.600	1.000	1.200	1.450	3.500
	Cholesterol Ratio	157	4.305	1.361	1.200	3.000	4.000	5.000	8.830
	Triglyceride	143	2.410	1.446	0.700	1.435	2.000	2.990	7.910
	Triglyceride/HDL Ratio	80	2.204	1.673	0.400	1.087	1.714	2.754	8.778
	Systolic	171	143.503	15.476	114.000	132.000	142.000	152.000	223.000
	Diastolic	171	84.164	10.321	65.000	78.000	80.000	90.000	122.000
	TSH	26	2.400	1.398	0.020	1.202	2.375	3.220	5.050
	T4	25	15.004	2.794	9.700	13.400	14.500	16.000	22.900
	HB	22	145.136	8.741	123.000	140.250	146.000	150.500	159.000
	HMCT	22	0.439	0.026	0.380	0.422	0.440	0.450	0.500
	months	256	0.000	0.000	0.000	0.000	0.000	0.000	0.000
1	Gender	256	0.590	0.493	0.000	0.000	1.000	1.000	1.000
	Age	256	63.424	12.387	23.167	54.750	62.750	73.500	91.500
	Height	75	1.706	0.092	1.473	1.625	1.720	1.770	1.900
	Weight	251	87.070	17.352	51.000	75.000	84.400	97.100	140.000
	BMI	65	30.356	5.923	19.240	27.040	29.270	32.470	53.620
	Diabetes	256	0.719	0.867	0.000	0.000	0.000	2.000	2.000
	HbA1c/ mmol/mol	201	45.925	9.319	32.000	40.000	43.000	50.000	84.000
	Gamma-G.T Level/ U/L	137	44.661	39.050	7.000	22.000	32.000	52.000	260.000
	Serum Cholesterol	174	4.892	1.247	2.400	4.025	4.700	5.700	8.800
	HDL Cholesterol	189	1.413	0.542	0.700	1.090	1.340	1.610	4.900
	Cholesterol Ratio	156	3.757	1.074	1.350	3.000	3.775	4.420	7.000
	Triglyceride	132	1.523	0.888	0.560	0.930	1.300	1.795	6.200
	Triglyceride/HDL Ratio	71	1.222	0.936	0.150	0.664	0.964	1.399	5.778
	Systolic	170	132.100	11.021	108.000	125.000	132.000	139.500	170.000
	Diastolic	170	77.794	7.570	54.000	71.250	78.000	82.000	110.000
	TSH	26	1.919	1.057	0.140	0.855	1.975	2.720	3.790
	T4	25	15.480	2.239	11.400	13.900	15.200	16.200	20.800
	HB	22	141.636	10.513	115.000	137.000	142.000	146.750	161.000
	HMCT	22	0.421	0.032	0.330	0.410	0.425	0.440	0.480
	months	256	22.199	17.456	1.000	8.000	19.000	32.000	84.000

Correlation Heatmap



I-Randomization Algorithm

(i.e., self-randomization by combining *subsampling* and *permutation test*)



Motivation: Using **two observations** of each subject (i.e. 2*256 samples) would violate the *Stable Unit Treatment Value Assumption*.

Idea: *Stage 1*--Using ***subsampling*** to mimic a randomized experiment by randomly pick either one of the two observations for each subject (i.e., 2^256 subsamples).
Repeat the subsampling M times and take average of ATEs.

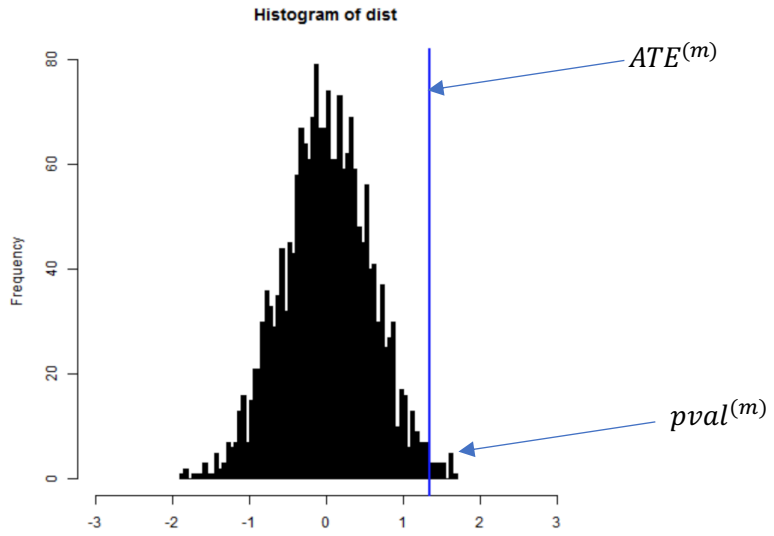
Stage 2--For each subsampling, using ***permutation test*** to evaluate the significance of ATE (e.g., obtaining p-values).

Stage 3--***Online FDR control*** for multiple testing (e.g., LORD algorithm by *Javanmard and Montanari, 2018*).

I-Randomization

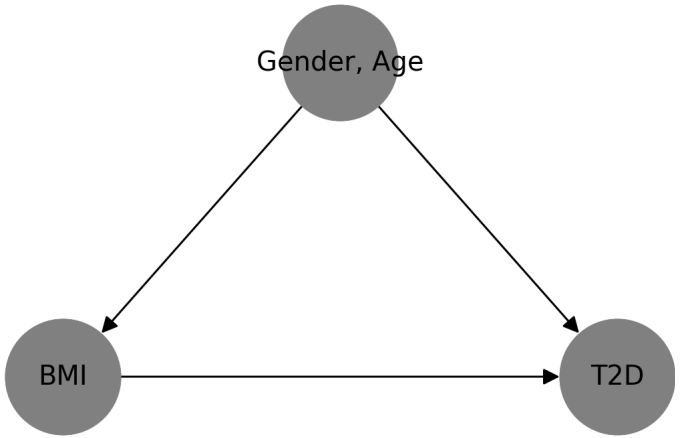
Pseudo Code

1. Start from a matrix with $2N$ rows where each data point is identified by the patient ID and the state (0 for the initial date and 1 for final date).
2. For m in 1 to M:
   (a) Sample a binary vector of length N; the index if the patient's ID and the value is the state (sample without replacement).
   (b) Select the corresponding subsample m.
   (c) Calculate $ATE^{(m)}$.
   (d) For s in 1 to S:
      - Shuffle the vector of treatment.
      - Calculate $ATE^{(m,s)}$ for the shuffle s of the treatment from subsample m.
   (e) Calculate the p-value $pval^{(m)}$ for the one-tailed test for the null hypothesis of no treatment effect, estimated by $\frac{1}{S}\sum_{s=1}^{S} \mathbb{1}_{ATE^{(m,s)}>(\text{resp. } s<)ATE^{(m)}}$.
3. Calculate the averaged $ATE = \frac{1}{M}\sum_{m=1}^{M} ATE^{(m)}$ and averaged p-value $= \frac{1}{M}\sum_{m=1}^{M} pval^{(m)}$.



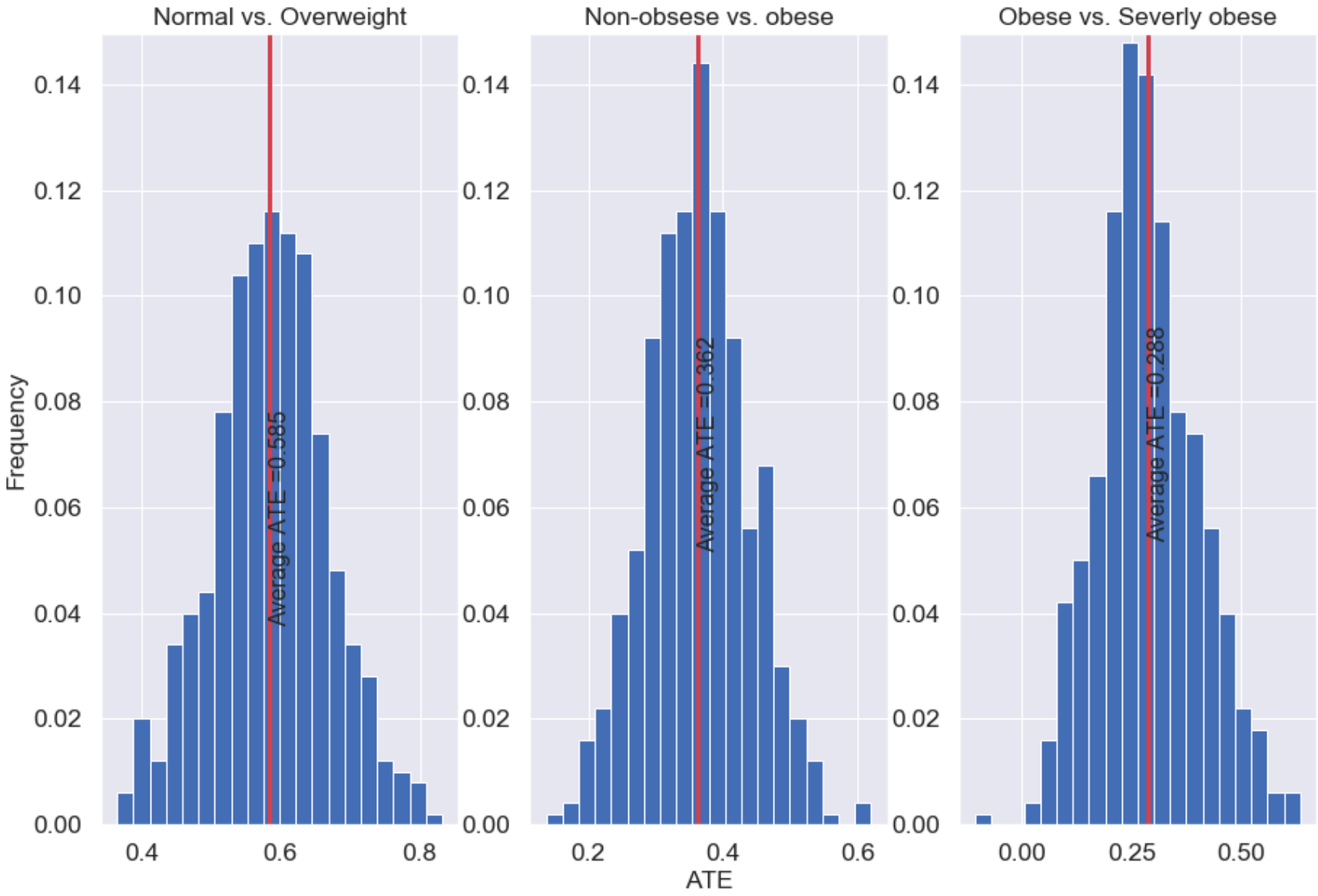
Values of $ATE^{(m,s)}, s = 1, \dots, S$

Experiment Result

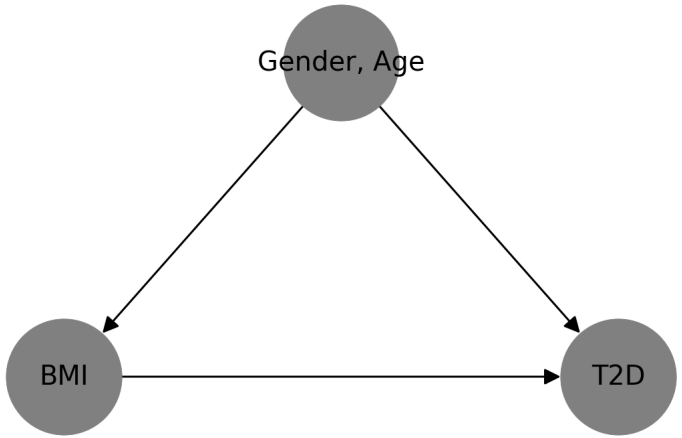


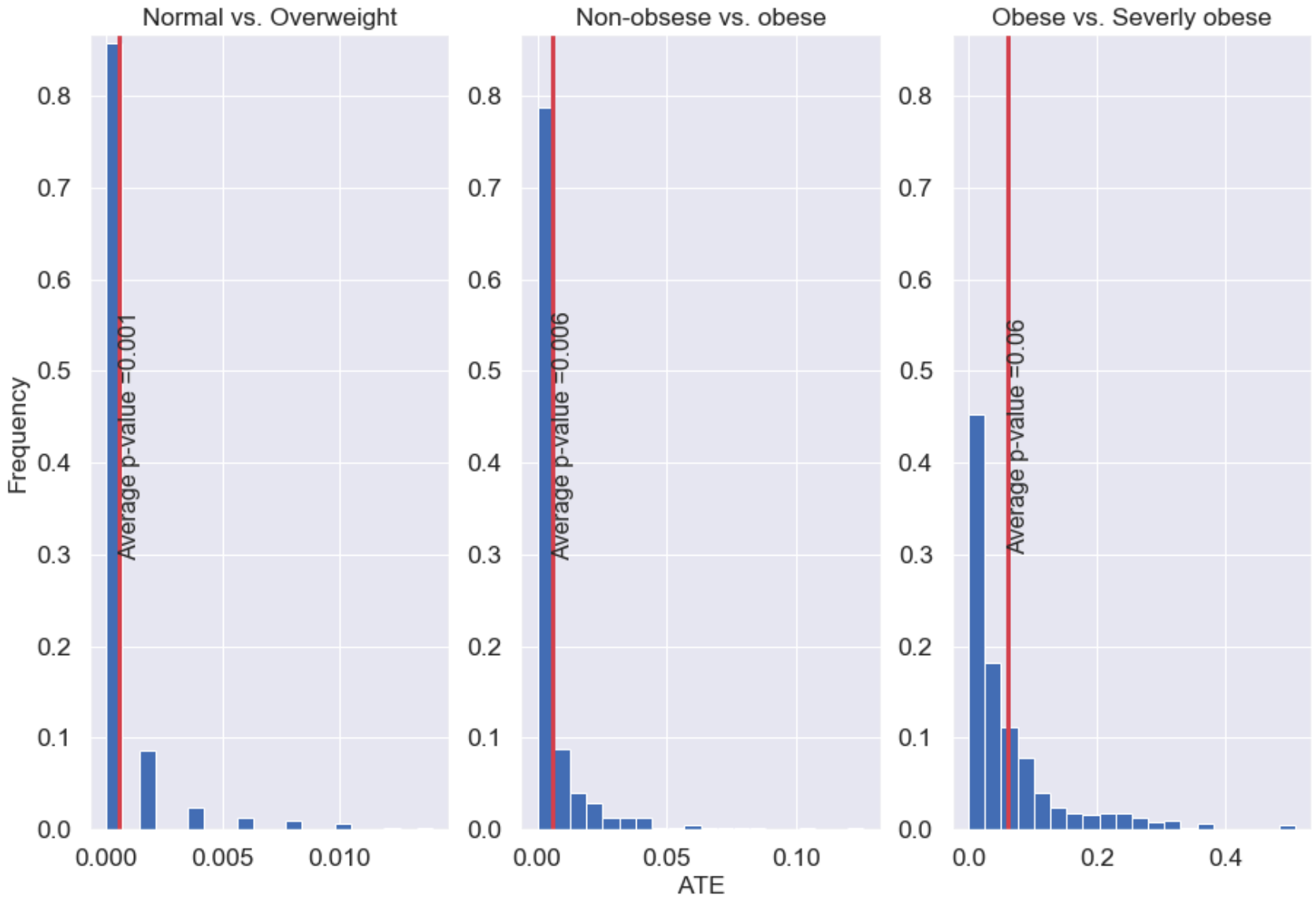
$$\text{BMI} = \begin{cases} \text{Normal } (< 25) \\ \text{Overweight } [25, 30) \\ \text{Obese } [30, 35] \\ \text{Severe Obese } > 35 \end{cases}$$

$$\text{T2D} = \begin{cases} \text{Non-Diabetic} \\ \text{Pre-Diabetes} \\ \text{Diabetes} \end{cases}$$



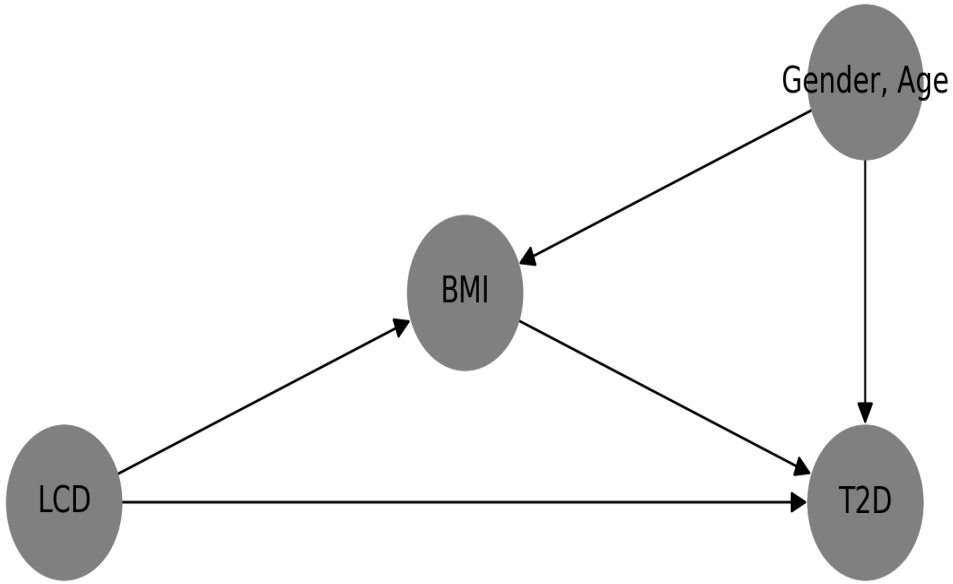
Significance Test





Q2: What causes obesity and type 2 diabetes?

Is obesity bad for type 2 diabetes?

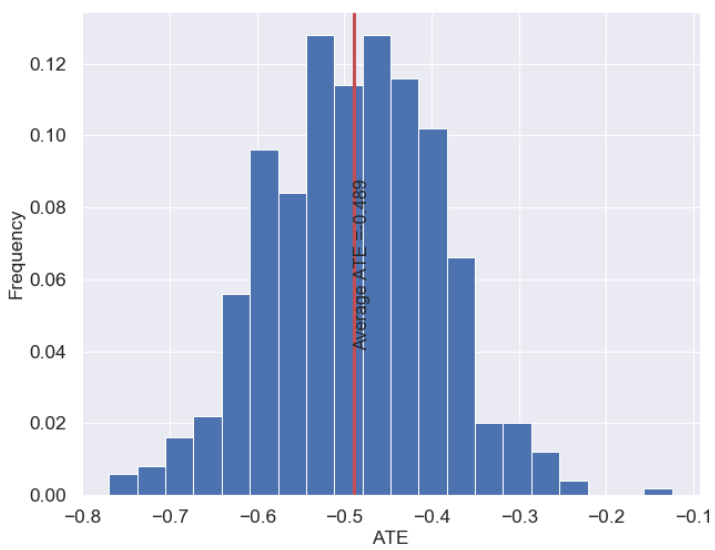


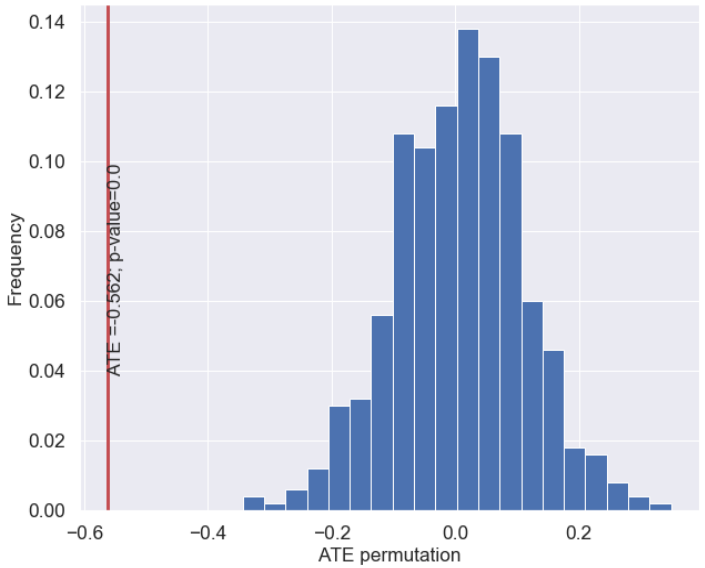
- BMI is a mediator between low-carbs diet and type-2 diabetes
- To estimate the effect of LCD on T2D we can:
  - Compute ATE without controlling for anything (BMI is a collider)
  - Control for all of BMI, Gender, and Age

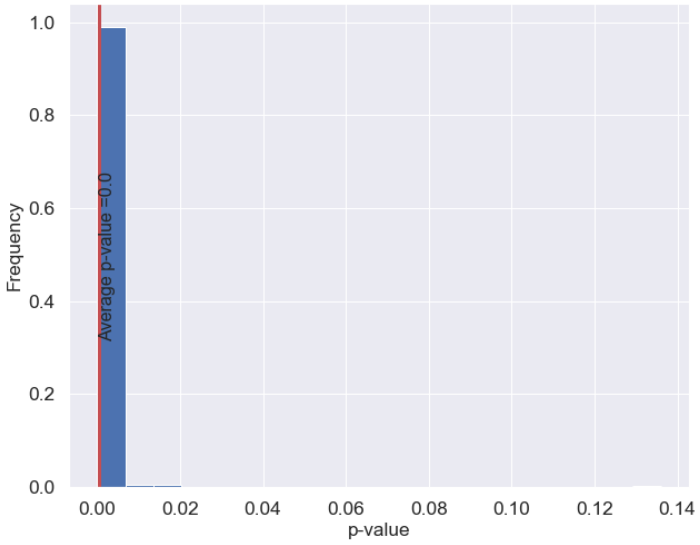
That's what we want to do

Yes, obesity is bad for type 2 diabetes … (counterfactual)

Distribution of the ATE from 500 subsamples (left), from one PT (middle), and of the p-value from 500 PT (right)



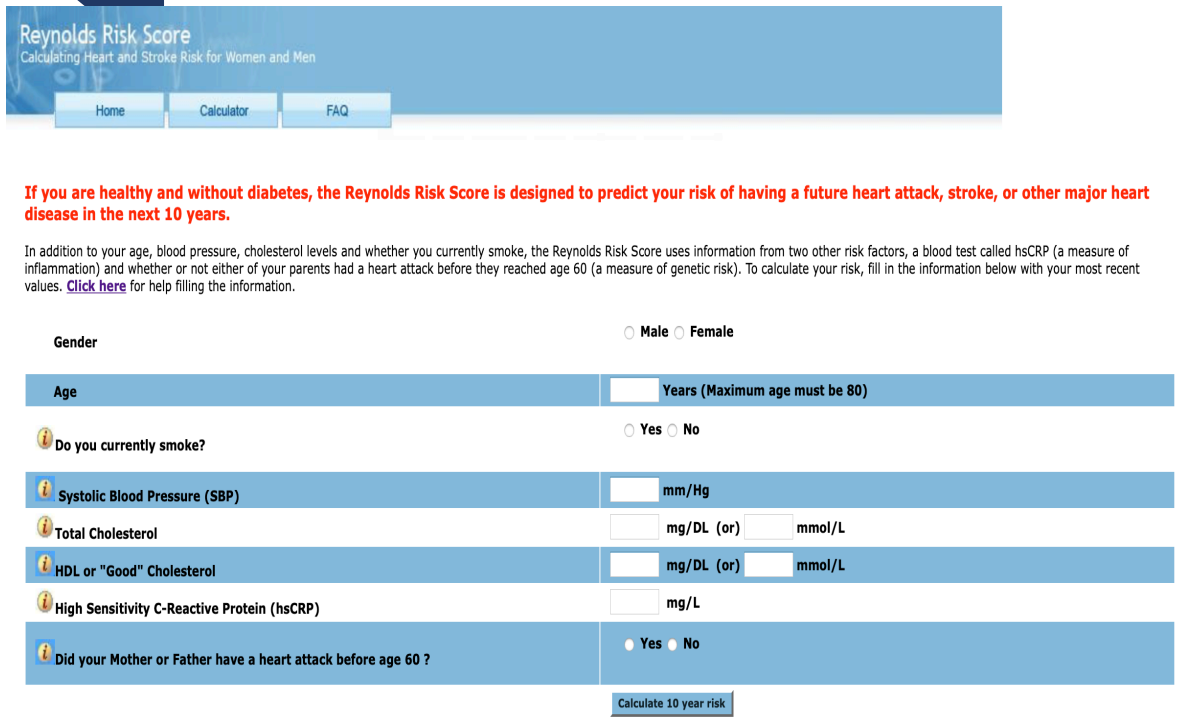




Q3: Does obesity cause high CVD risk?

Reynolds Risk Score for CVD

P.M. Ridker, et al., 2007, JAMA



Assumptions due to limited data

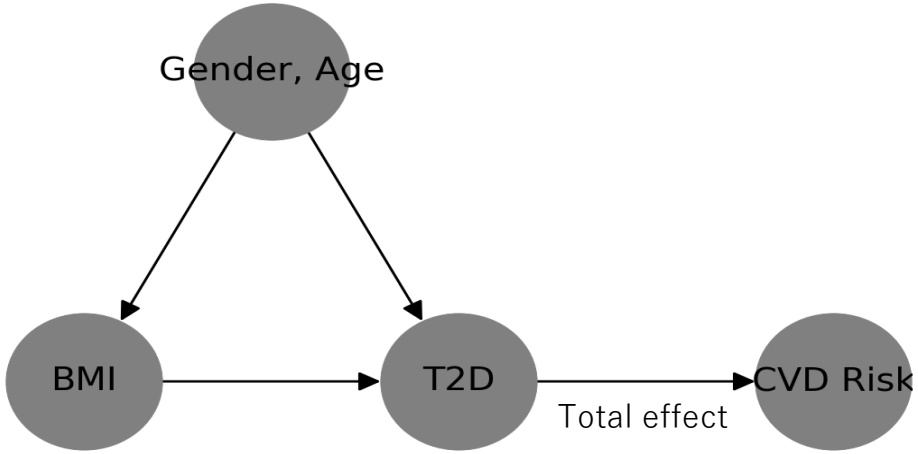
- Assume all patients are non-smokers
- Ignore hsCRP
- Ignore information about family history

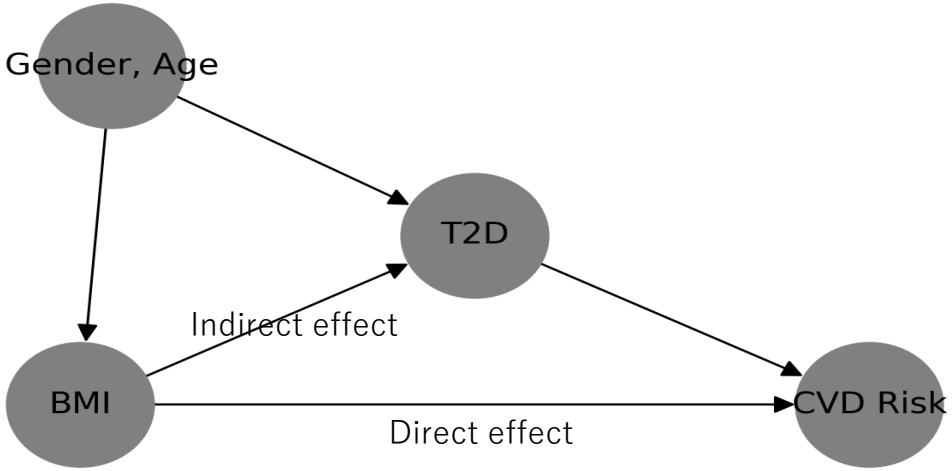
Reynolds Risk Score (RRS) does not include diabetes information

We can also use Blood Pressure (BP) or Cholesterol (HDL and Total) as the outcome

Impact of obesity on CVD Risk

Impact of Obesity (BMI) on CVD Risk





- BMI diet impacts CVD Risk through T2D
  - T2D is a mediator and no direct path from LCD to CVD Risk
  - We control for Gender, and Age

CVD Risk | BMI; Gender, Age

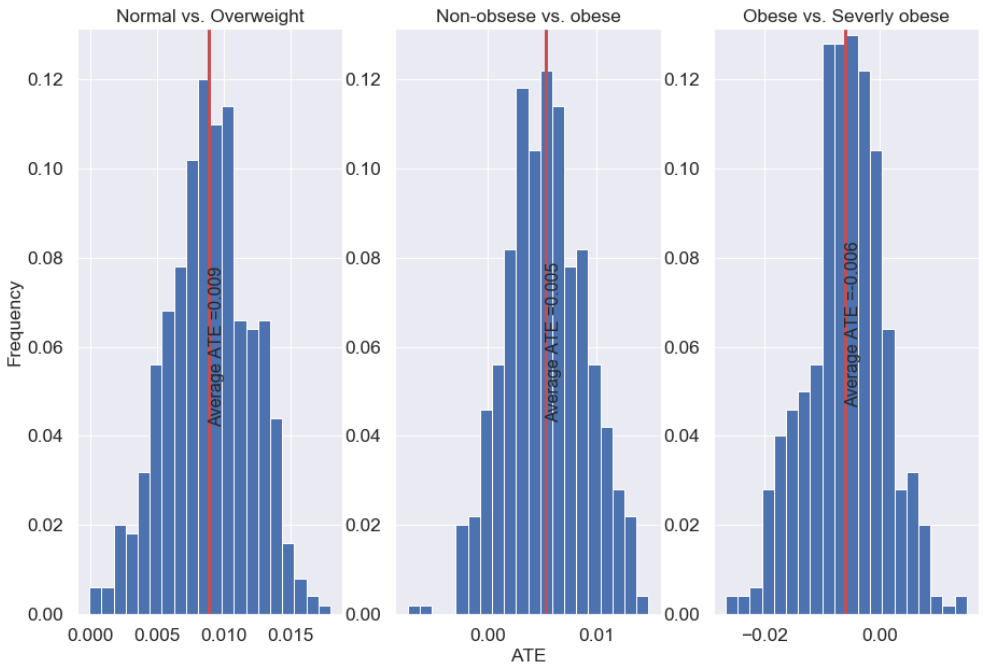
- BMI diet impacts CVD Risk through T2D
  - T2D is a mediator with a direct path from LCD to CVD Risk
  - We control for all of Gender, Age, and T2D

CVD Risk | BMI; Gender, Age, T2D

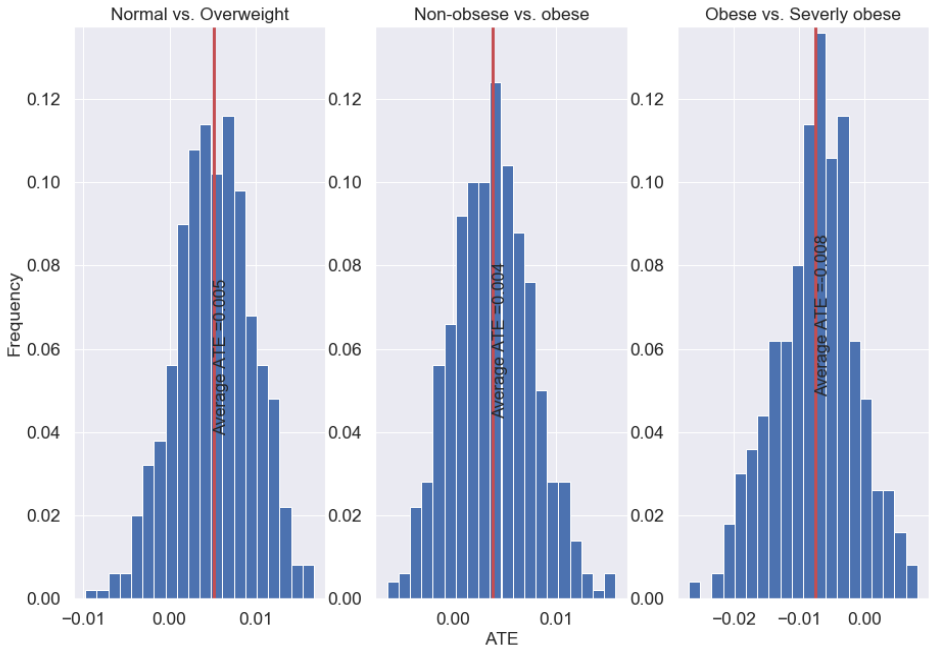
Impact of obesity on CVD Risk

Distribution of subsamples

RRS | BMI; Gender, Age



RRS | BMI; Gender, Age, T2D



Indirect causal effect = Total effect – Direct effect

0.009 – 0.005 = 0.004 (Normal vs. Overweight)

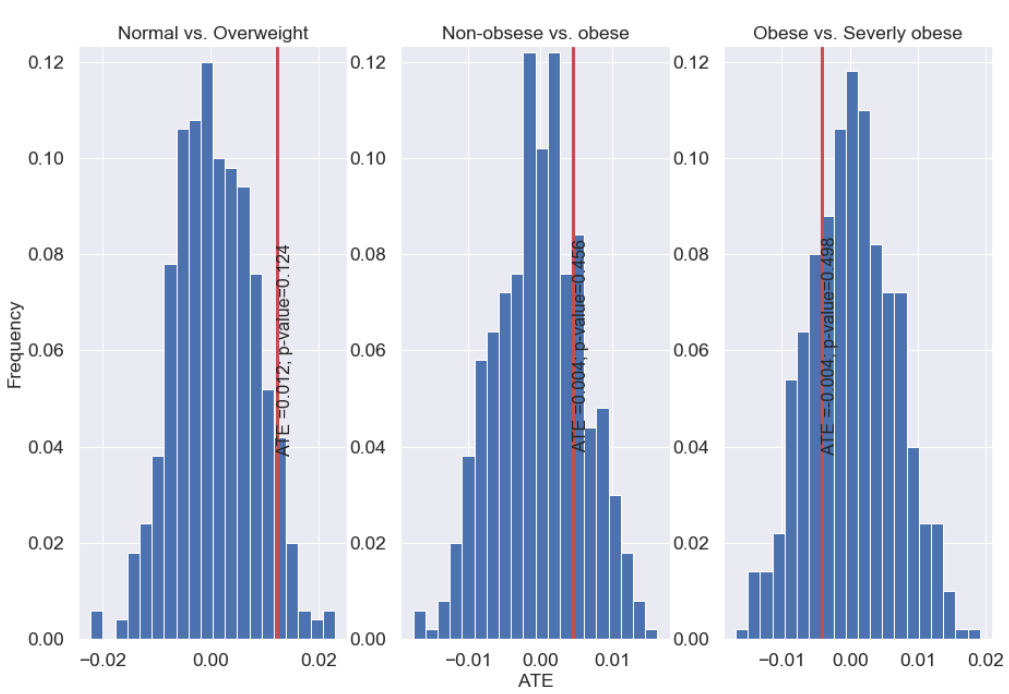
0.005 – 0.004 = 0.001 (Non-obese vs. Obese)

0.006 – 0.008 = -0.002 (Obese vs. Severely obese)

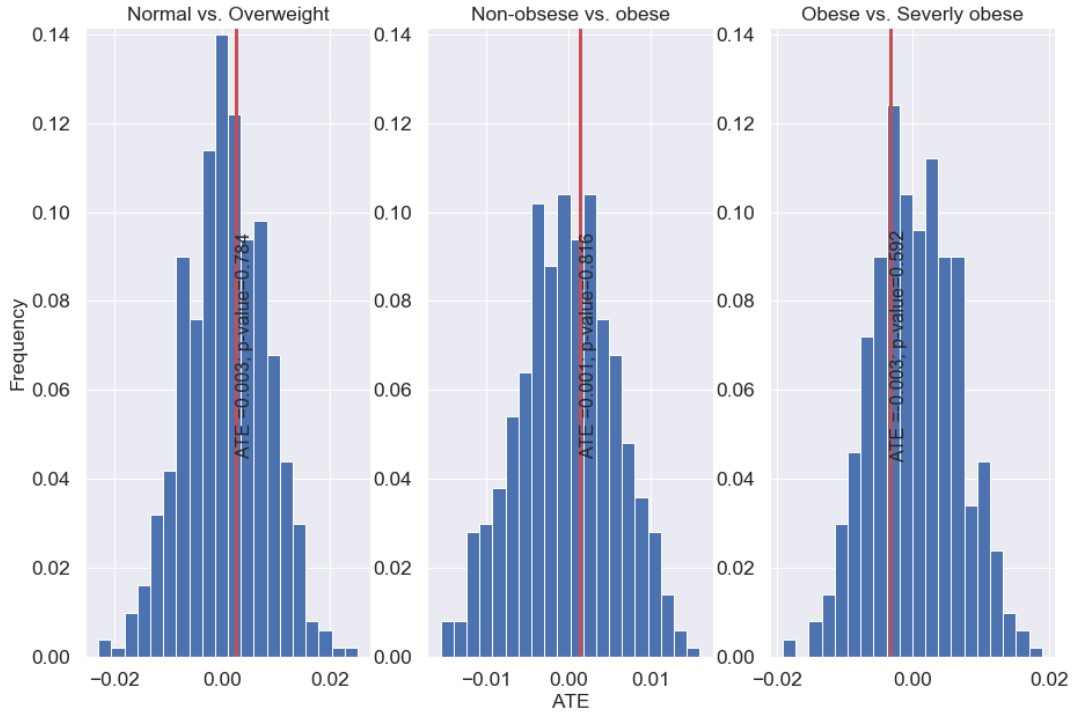
Impact of obesity on CVD Risk

Distribution of one permutation test

RRS | BMI; Gender, Age



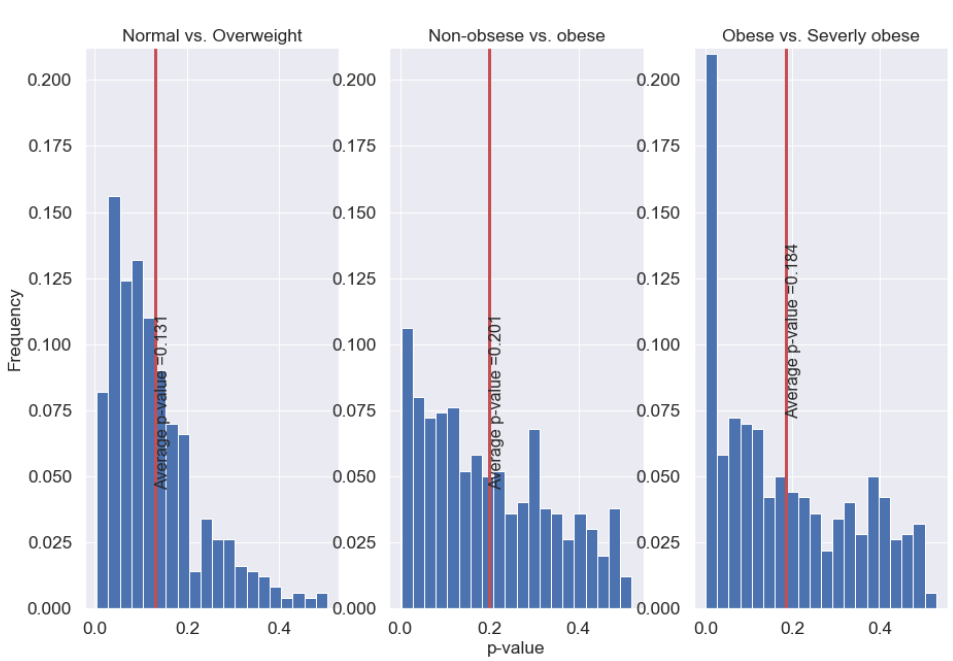
RRS | BMI; Gender, Age, T2D



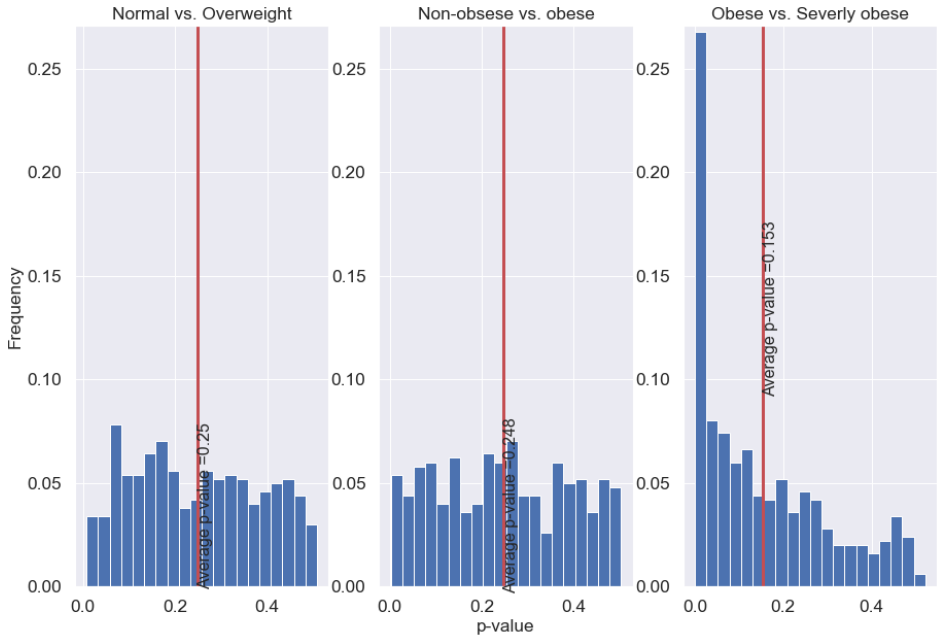
Impact of obesity on CVD Risk

Distribution of p-values

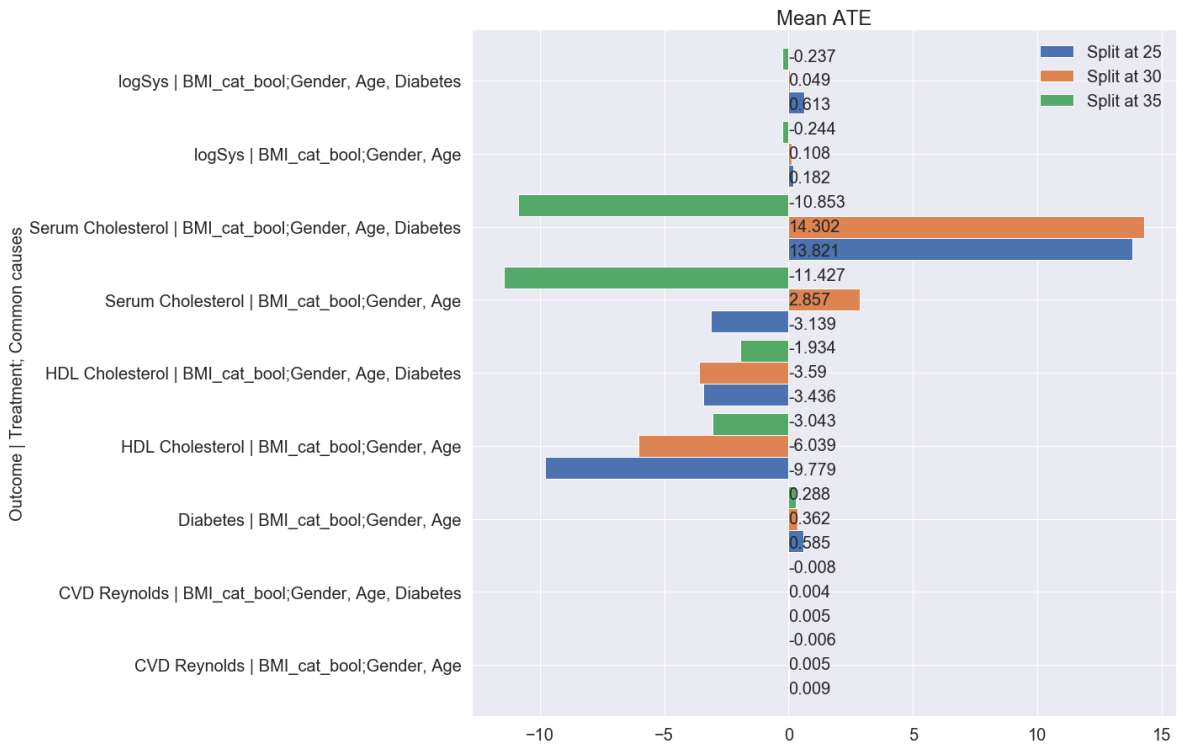
RRS | BMI; Gender, Age

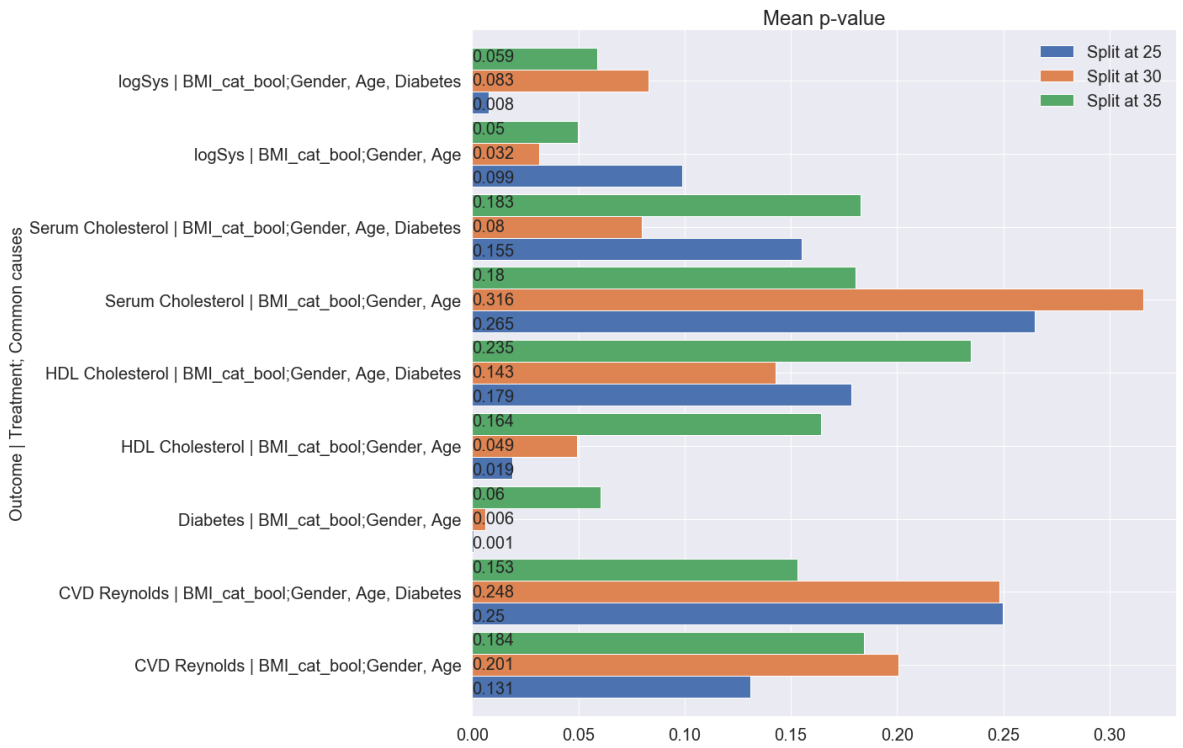


RRS | BMI; Gender, Age, T2D

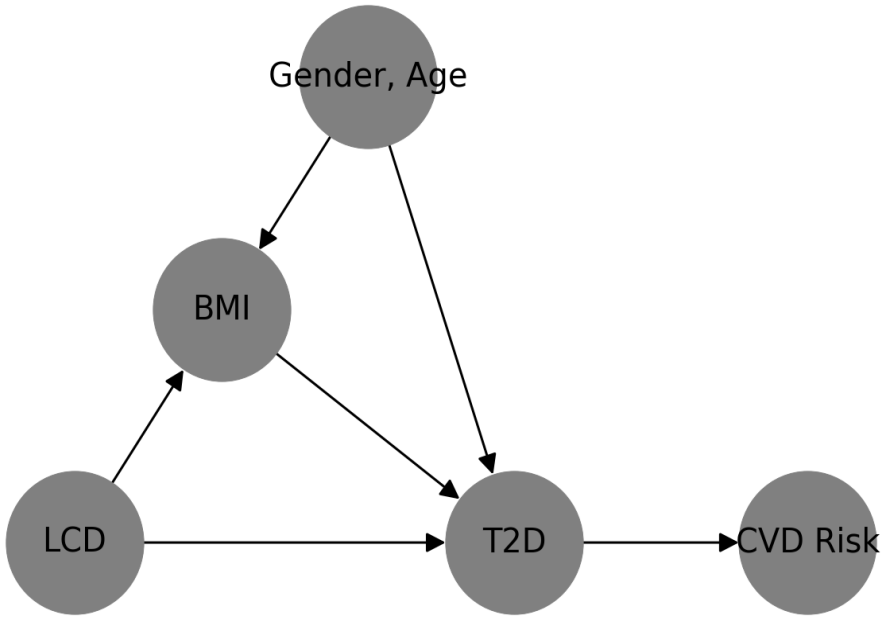


Summary of the Effect of Obesity





Low-Carb is promising for lowering CVD risk !



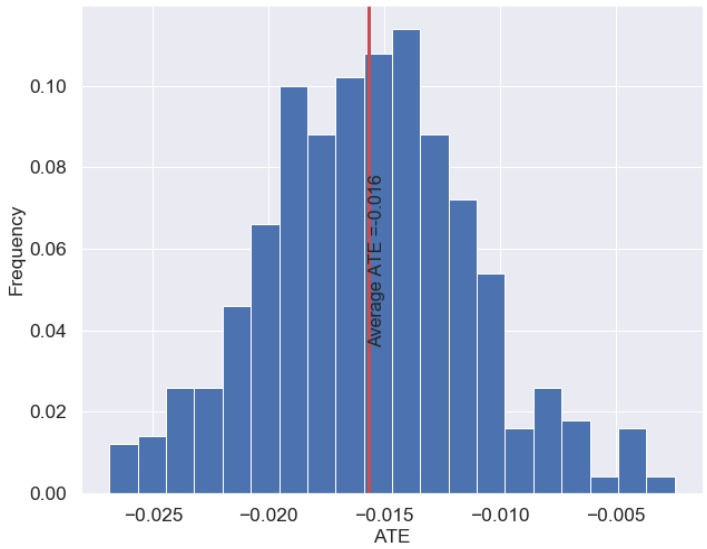
RRS | LCD; Gender, Age, BMI

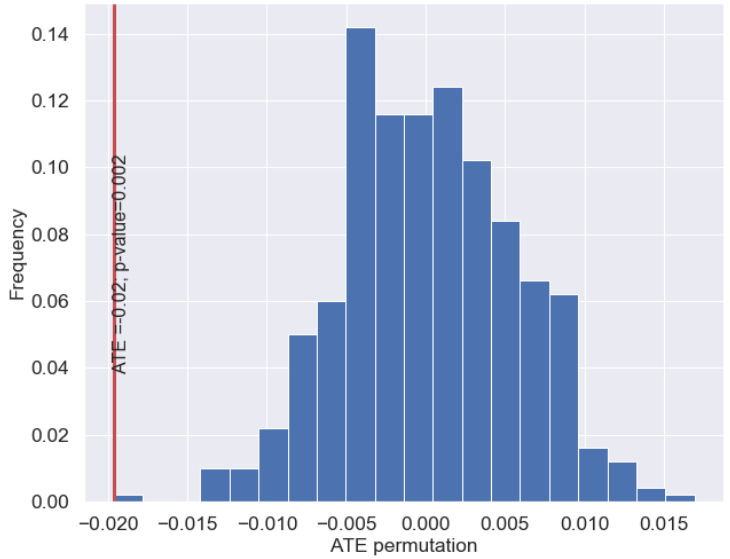
Impact of low-carb diet on CVD Risk

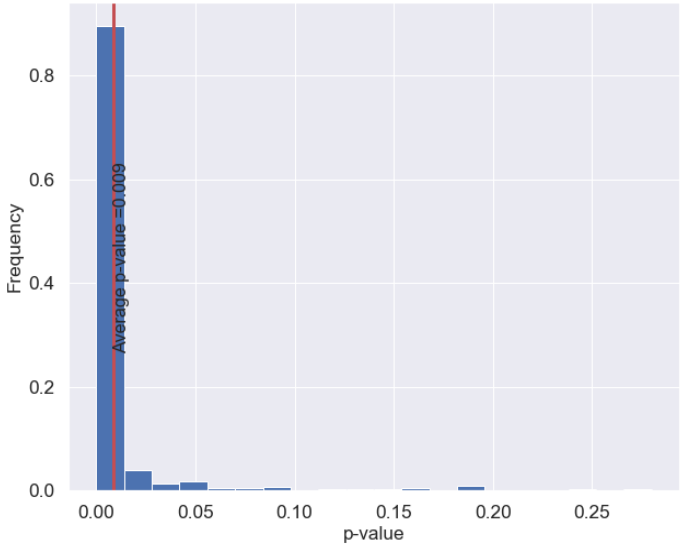
- Low-carb diet impacts CVD Risk through T2D
  - T2D is a mediator
  - No direct path from LCD to CVD Risk
  - Control for BMI, Gender, and Age

Impact of Low-Carb Diet on CVD risk

RRS | LCD; Gender, Age, BMI

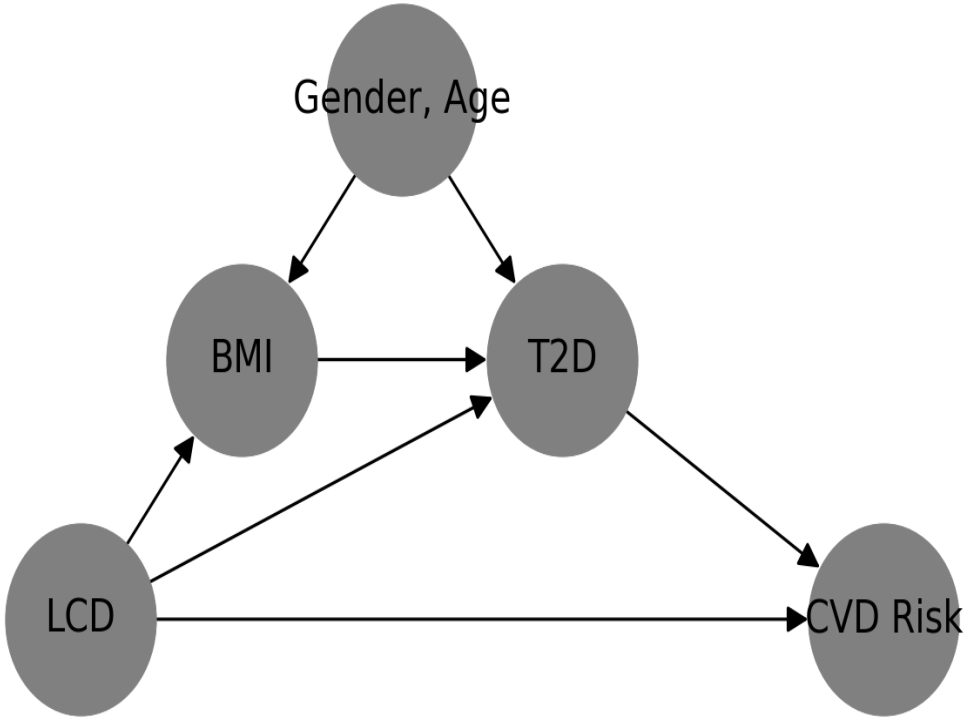






Distribution of the ATE from 500 subsamples (left), from one PT (middle), and of the p-value from 500 PT (right)

Low-Carb is promising for lowering CVD risk !



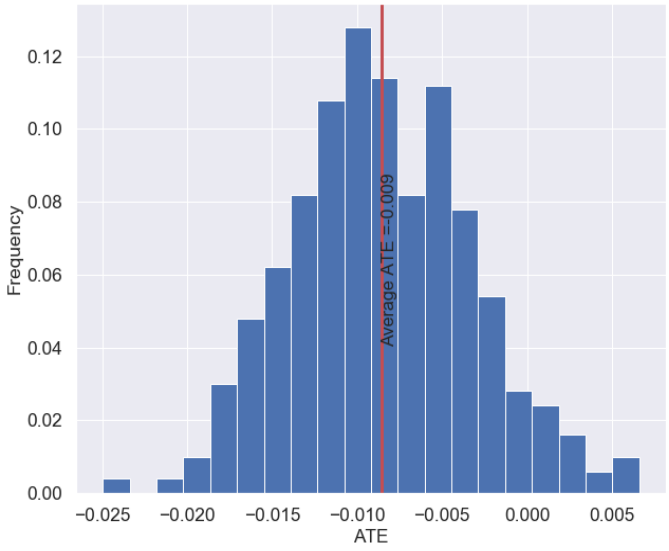
RRS | LCD; Gender, Age, BMI, T2D

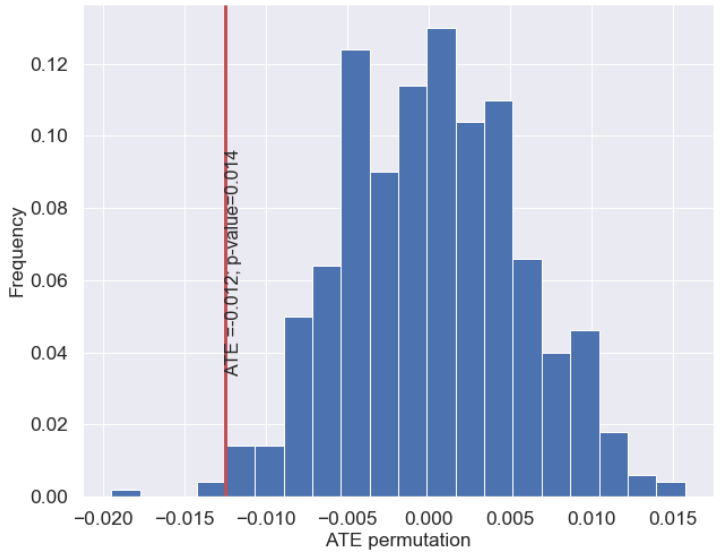
Impact of low-carb diet on CVD Risk

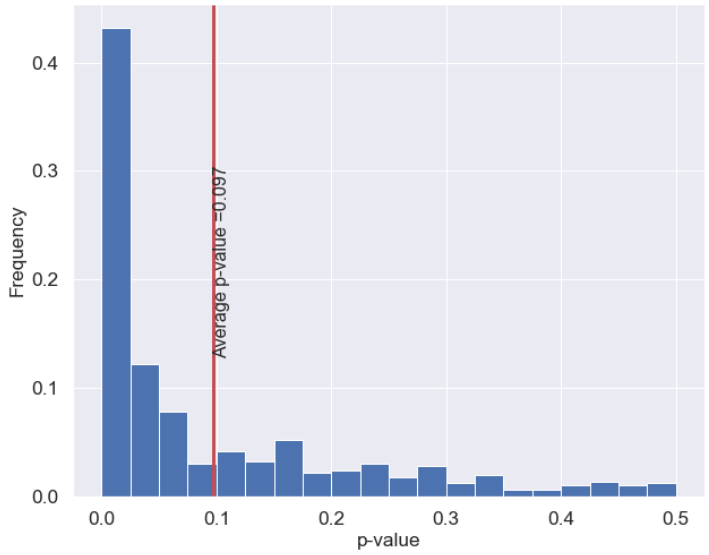
- Low-carb diet impacts CVD Risk both directly and indirectly through
  - T2D is a mediator
  - There is also a direct path from LCD to CVD Risk
  - We control for Gender, Age, and T2D

Low-Carb is promising for lowering CVD risk !

RRS | LCD; Gender, Age, BMI, T2D

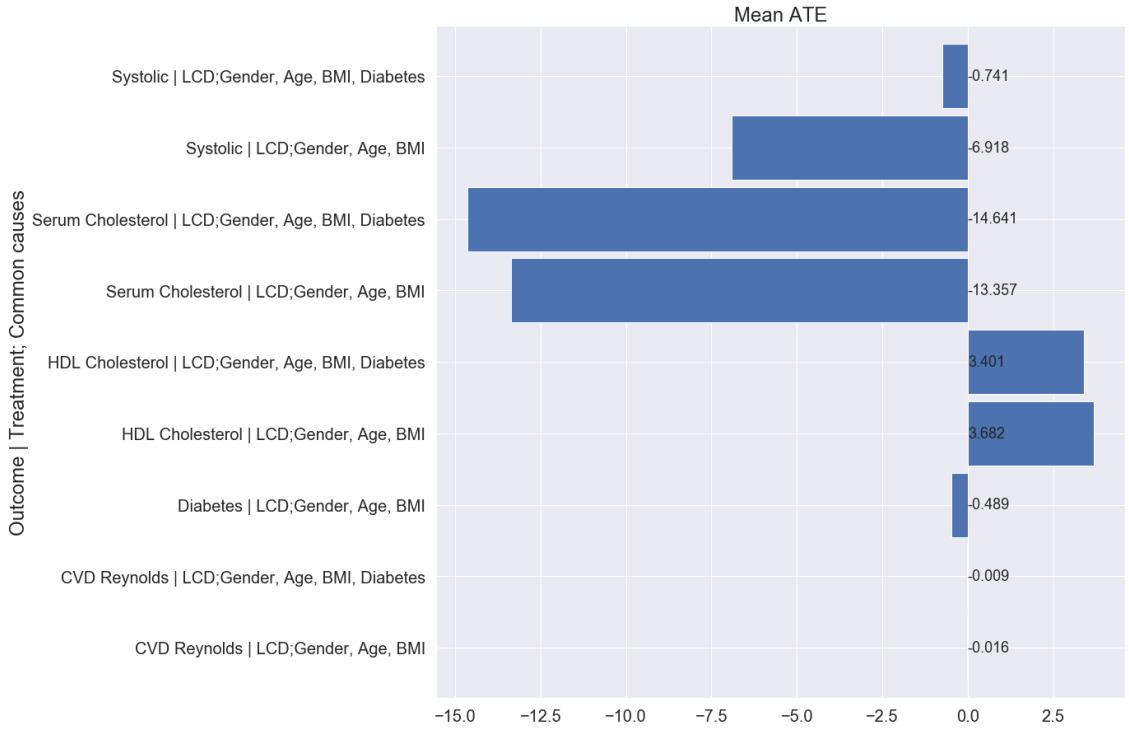


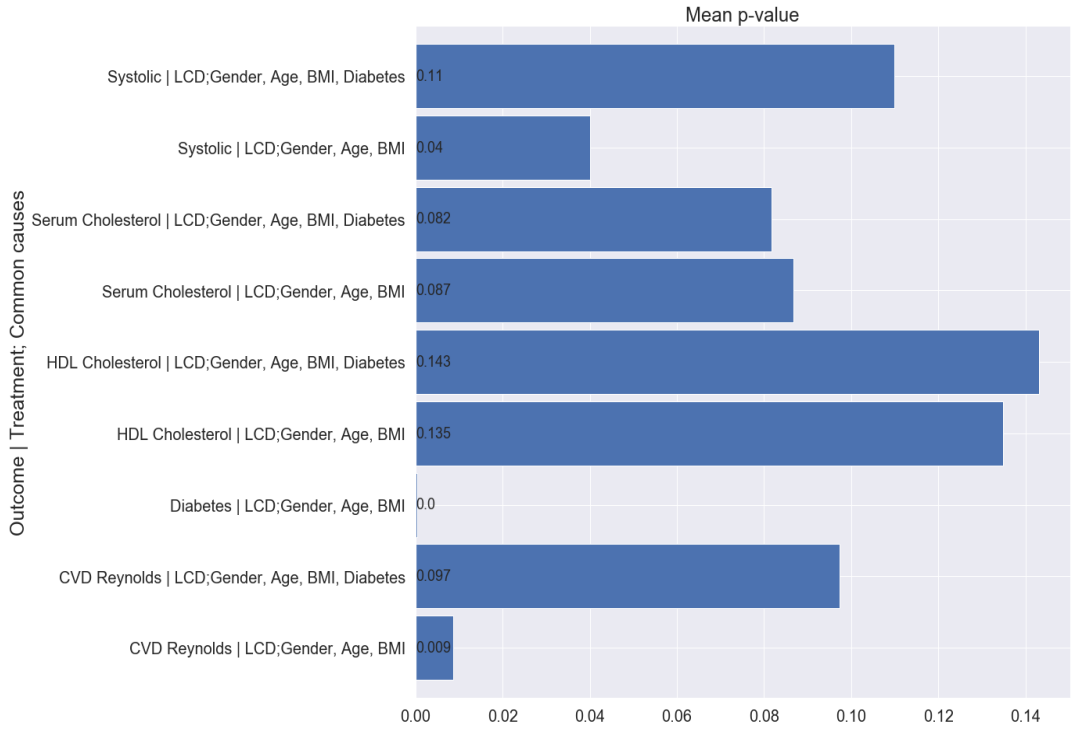




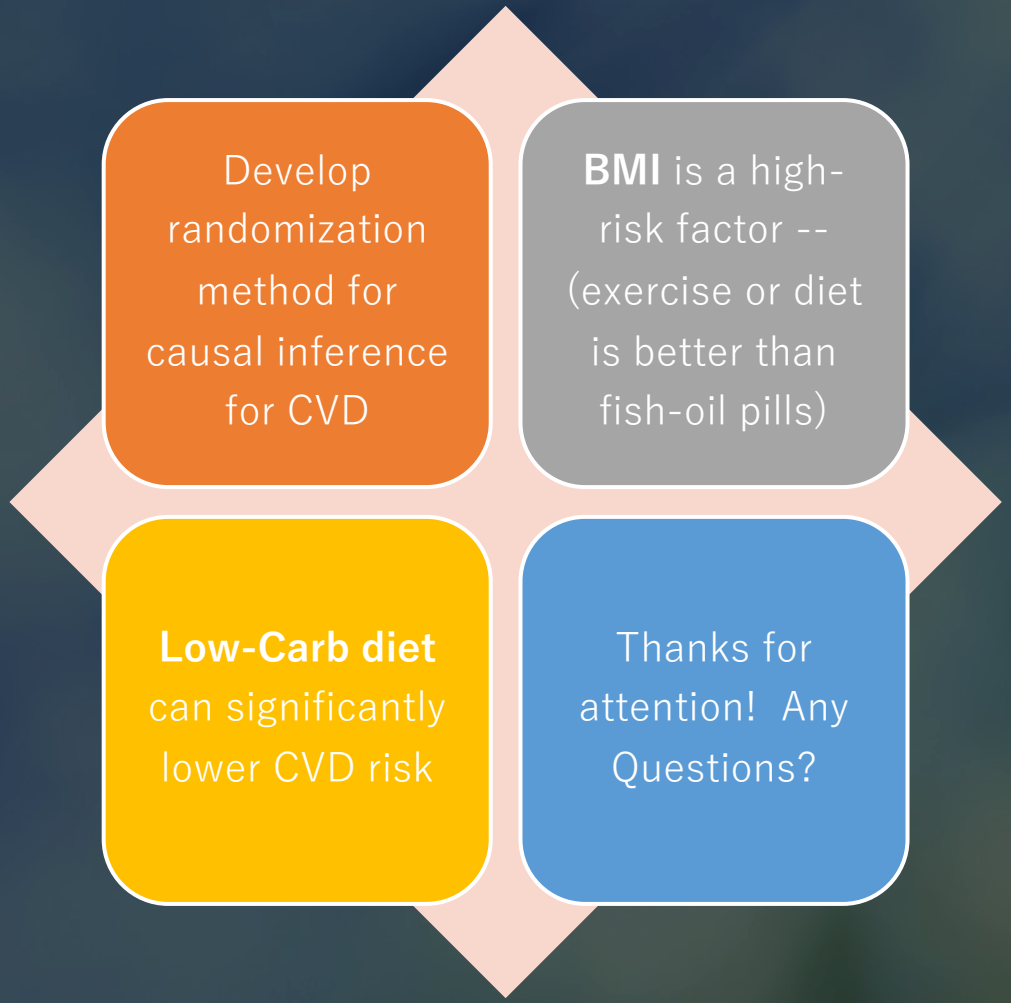
Distribution of the ATE from 500 subsamples (left), from one PT (middle),
and of the p-value from 500 PT (right)

Low-Carb is promising for lowering CVD risk and Diabetes





Conclusion


Develop randomization method for causal inference for CVD
BMI is a high-risk factor -- (exercise or diet is better than fish-oil pills)
Low-Carb diet can significantly lower CVD risk
Thanks for attention! Any Questions?