Causal Assumptions

ehsan.karim@ubc.ca

Oct 10, 2021

SPPH 504/007

Reference

- Hernán MA, Robins JM (2020). Causal Inference: What If. Boca Raton: Chapman & Hall/CRC (link)

Notations

Outcome

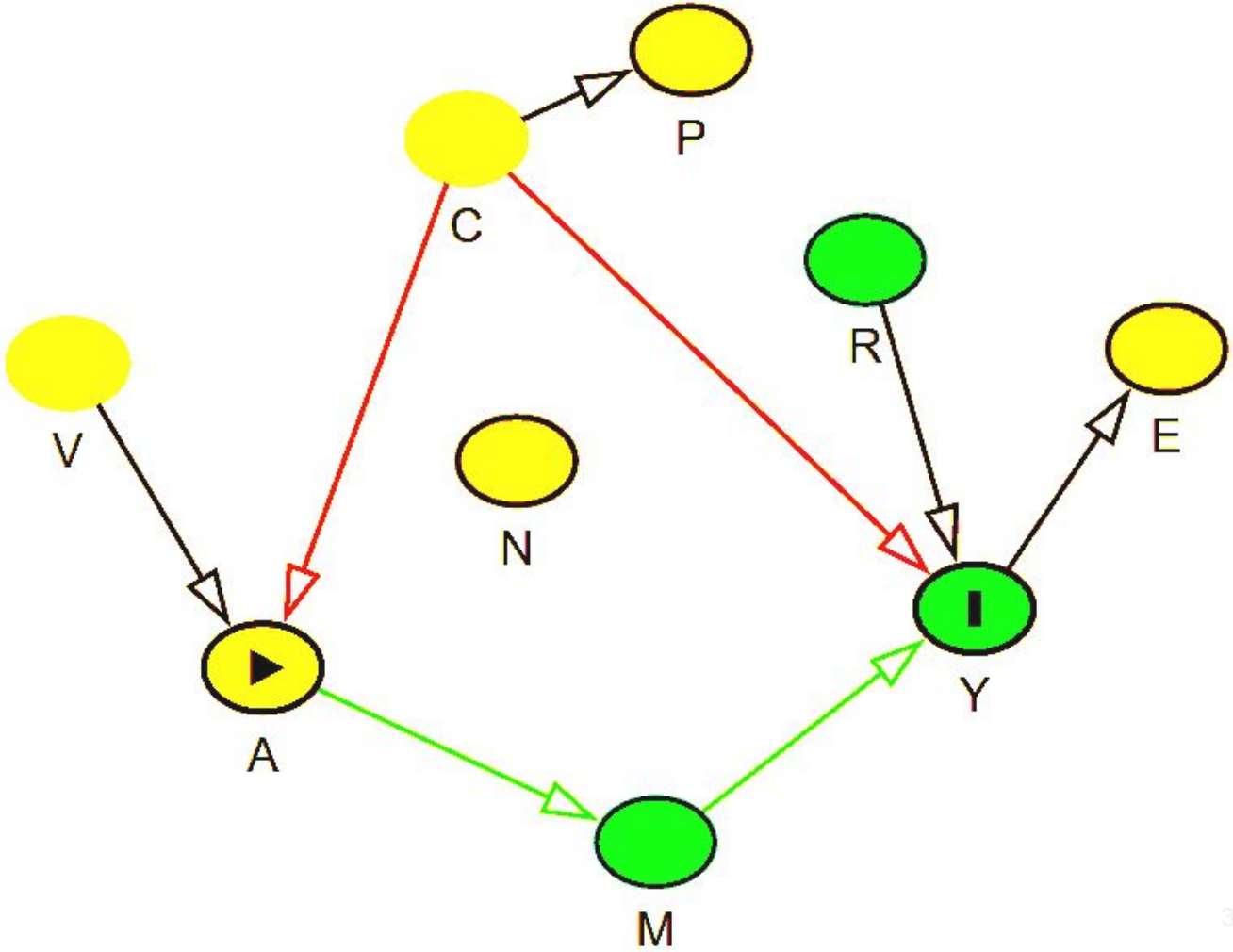
Treatment

Confounder

Risk factors

Effect

Noise



RCT

- Treatments are randomized.
- Objective is to estimate treatment effect.
- If enough sample size
  - Confounding should not be an issue
    - Observed
    - unobserved

How to estimate unbiased treatment effect from an RCT? Y = outcome, A = treatment, C = confounder, R = Pure risk factors for outcome, V = Determinants of treatment assignment

Y ~ Indicator for the groups determined by randomization

- Y ~ A
- Y ~ A + C
- Y ~ A + C + R
- Y ~ A + C + R + V
- Y ~ A + R
- Y ~ A + V

What changes when randomization is not there?

- Need to think why RCT was working
- If we can meet the same conditions, observational data analysis may have some merit
- What is RCT achieving?

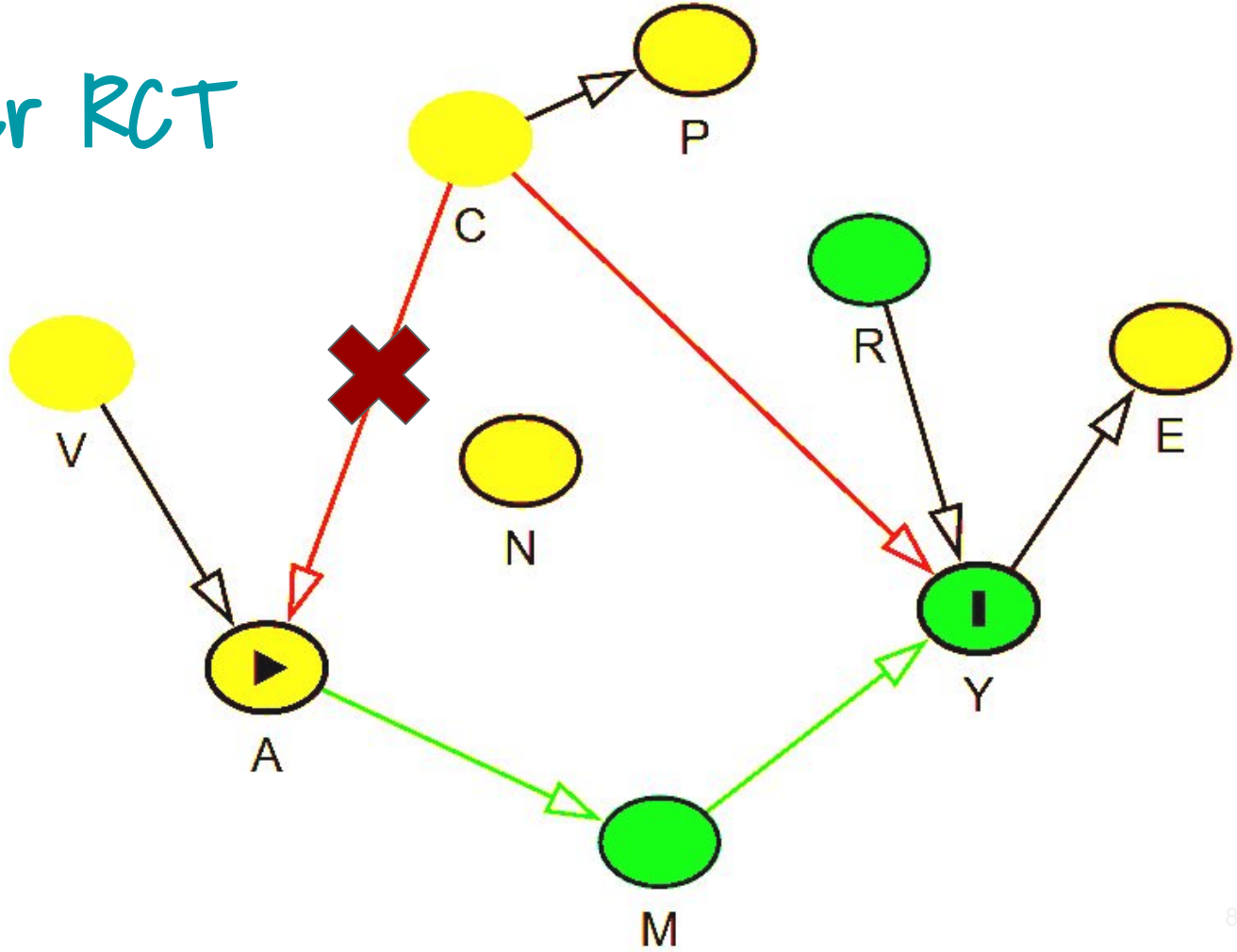
Table 1

RHC

Table 1.—Characteristics of 5735 Critically Ill Patients*

Variable	No RHC (n=3551)	RHC (n=2184)
Age range, y†		
<50	884 (25)	540 (25)
50 to <60	546 (16)	371 (17)
60 to <70	812 (23)	577 (26)
70 to <80	809 (23)	529 (24)
>80	500 (14)	167 (8)
Sex†		
Male	1914 (54)	1218 (59)
Female	1637 (46)	906 (41)
Race		
White	2753 (78)	1707 (78)
Black	585 (17)	335 (15)
Other	213 (5)	142 (7)

Notations under RCT



What changes when randomization is not there?

- Need additional considerations
  - Identifiability conditions
    - P(A|L) depends on measured L
      - No unmeasured confounding, exchangeability
      - Y(a) independent of A | L
    - A well-defined?
      - Causal consistency
    - P(A|L) > 0
      - Positivity

Exchangeability

- John takes rosuvastatin ($A = 1$) and his cholesterol level = 200
- Jim do not take rosuvastatin ($A = 0$) and his cholesterol level = 250
- If Jim took rosuvastatin ($A = 1$), and if his cholesterol level was same as John (200), then we say that Jim and John are exchangeable.

Conditional Exchangeability

Exchangeable within same sex: Y(a) independent of A | Sex

Name	Y(1): outcome when takes tx	Y(0): outcome when does not take tx	Sex
John	200	250	Male
Jim	200	250	Male
Kate	150	200	Female
Hilda	150	200	Female

Conditional Exchangeability

Exchangeable within same sex and age group: Y(a) independent of A | (sex, age)

Name	Y(1): outcome when takes tx	Y(0): outcome when does not take tx	Sex	Age
John	200	250	Male	20
Jim	200	250	Male	20
Kate	150	200	Female	20
Hilda	150	200	Female	20
Joseph	400	500	Male	90
Jack	400	500	Male	90
Anna	300	400	Female	90
Melissa	300	400	Female	90

Observed data

Exchangeable within same sex and age group: Y(a) independent of A | (sex, age)

Y~ A + sex + age

Name	Y(1): outcome when takes tx	Y(0): outcome when does not take tx	Sex	Age
John		250	Male	20
Jim	200		Male	20
Kate	150		Female	20
Hilda		200	Female	20
Joseph	400		Male	90
Jack		500	Male	90
Anna	300		Female	90
Melissa		400	Female	90

Observed data

Given some data, how are you analyzing the data?
Assuming conditional exchangeability: we analyze
Y~ A + sex + age

Name	Y(1): outcome when takes tx	Y(0): outcome when does not take tx	Sex	Age	U
Subject 1		251	Male	20	?
Subject 2	199		Male	20	?
Subject 3	151		Female	20	?
Subject 4		210	Female	20	?
Subject 5	390		Male	90	?
Subject 6		480	Male	90	?
Subject 7	303		Female	90	?
Subject 8		401	Female	90	?

How to select covariates to meet conditional exchangeability?

Checking balance stratifying by exposure
Empirical selection (Stepwise regression) with A being outcome
Empirical selection (Stepwise regression) with Y being outcome
Subject area knowledge
Big data analytics
Modified disjunctive cause criterion
Automatic High-Dimensional “Proxy” Adjustment
Machine learning variable importance
Combining propensity score with empirical selection
Change-in-estimate

Positivity

$\Pr(A = a \mid L = l) > 0$

$\Pr(A=1 \mid \text{sex} = \text{male}) > 0$

$\Pr(A=1 \mid \text{sex} = \text{female}) > 0$

$\Pr(A=0 \mid \text{sex} = \text{male}) > 0$

$\Pr(A=0 \mid \text{sex} = \text{female}) > 0$

Can Positivity assumption be empirically verified from the data?

- Yes
- No
- Sometimes
- Practically impossible as all covariates can't be measured

Positivity

$Pr(A = a \mid L = l) > 0$

$Pr(A=1 \mid \text{eye color} = \text{black}) > 0$

$Pr(A=1 \mid \text{eye color} = \text{brown}) > 0$

$Pr(A=1 \mid \text{eye color} = \text{blue}) = 0$

- Structural
  - Male pregnancy
- Random
  - Not really 0, but it can happen due to small sample size
  - Zero-cell correction?

Eye color has anything to do with Y and A?

Positivity only required for L's that are relevant for conditional exchangeability.

https://ehsanx.shinyapps.io/project0/

Observed data

Name	Y(1): outcome when takes tx	Y(0): outcome when does not take tx	Sex	Age
John		250	Male	20
Jim	200		Male	20
Kate	150		Female	20
Hilda		200	Female	20
Joseph			Male	90
Jack		500	Male	90
Anna			Female	90
Melissa		400	Female	90

Causal Consistency

$Y(a) = Y$ for everyone receiving $A = a$

($A = 1$ == rosuvastatin 5 mg vs. $A = 0$ == no treatment)

- John's cholesterol level = 200 if he takes rosuvastatin 5 mg ($A = 1$)
- John's cholesterol level = 250 if he does not take rosuvastatin ($A = 0$)

John's $Y(A=1) = 200$

John's $Y(A=0) = 250$

Need to specify version: A = rosuvastatin 5 mg

Causal Consistency

Need to specify version: *A* = rosuvastatin 5 mg

We know often John breaks a 10 mg and takes one-half on 2 separate occasions. Often while breaking the tablet, the split is not exactly 5 mg. Could be 4.5 or 5.5 mg. Is that sufficiently well-defined? Is that meaningfully different? Realistic?

Treatment-variation irrelevance can be an approximation: two IFNbeta-1a products (Rebif and Avonex) and one IFNbeta-1b product (Betaferon)

We want to find out causal effect of overweight (A: BMI is 25.0 to <30) at age 50 on the risk of mortality (Y) by age 55 in British Columbia. Is A sufficiently well-defined?

No, A being BMI = 25.7 would be better defined.

I think so. It is practical.

No. This is ill-defined.

Assumptions related to Mediation Analysis

- General assumptions (mediator acts as an added exposure)
  - Conditional exchangeability
  - Positivity
  - Causal consistency
- Additional
  - Model specification (not specific to mediation; applies to total effect models as well)
  - No interaction between exposure and mediator

Assumption - 1

- L is sufficient to address confounding. No uncontrolled confounding in:
  - exposure-outcome associations
    - Y(A=a, M(a)) independent of A assignments given L
  - exposure-mediator associations
    - M(a) independent of A assignments given L
  - mediator-outcome associations
    - Y(A=a, M(a)) independent of M assignments given L
- One related idea is model-misspecification
  - Generally good to consider realistic/plausible interactions between
    - Exposure * covariate; or Mediator * covariate; or covariate * covariate

Assumptions - 2, 3 & 4

- Positivity
  - All exposure values have non-zero probability for any values of L
    - $P(A=a|L=l) > 0$ for all a and l
  - All mediator values have non-zero probability for any values of A & L
    - $P(M=m|A=a, L=l) > 0$ for all m, a and l
- Causal Consistency
  - Observed values are realistic
  - No multiple version of A or M
- No exposure-mediator interactions

Thanks!


ehsan.karim@ubc.ca


www.ehsankarim.com