Balance and SMD!

ehsan.karim@ubc.ca

Oct 7, 2020

SPPH 504/007

Illustrative example: Research question

Research Question: Whether or not adult patients with rheumatoid arthritis (RA) are at increased risk for heart attack (or myocardial infarction) in US.

Outcome (Y): heart attack (MI)

Exposure (A): rheumatoid arthritis (RA)

Comparison group: People without RA.

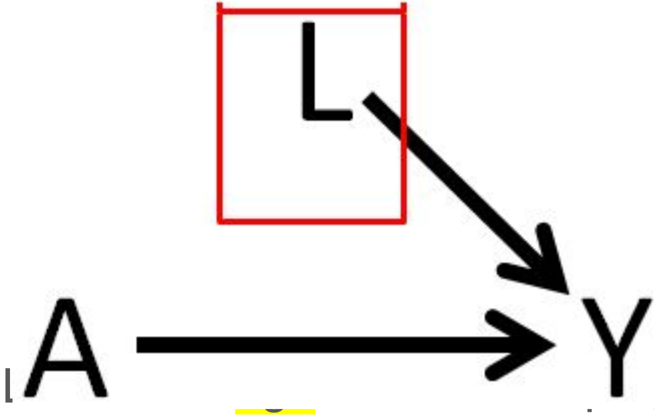
Exclusion criteria: Patients with

$$A \longrightarrow Y$$

Osteoarthritis or other arthritis, young subjects (age < 20).

RCT

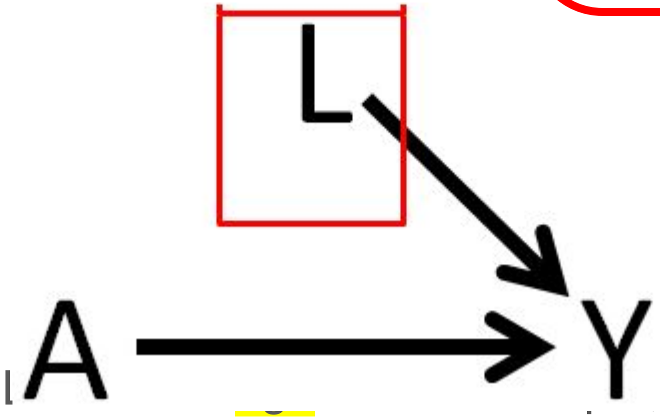
RCT



A = RA; Y = MI

RCT

RCT



	Non-arthritis	Rheumatoid
age (%)		
(0,50]	60 (30.9)	62 (32.0)
(50,70]	82 (42.3)	85 (43.8)
70+	52 (26.8)	47 (24.2)

age distribution balanced in 2 groups

A = RA; Y = MI

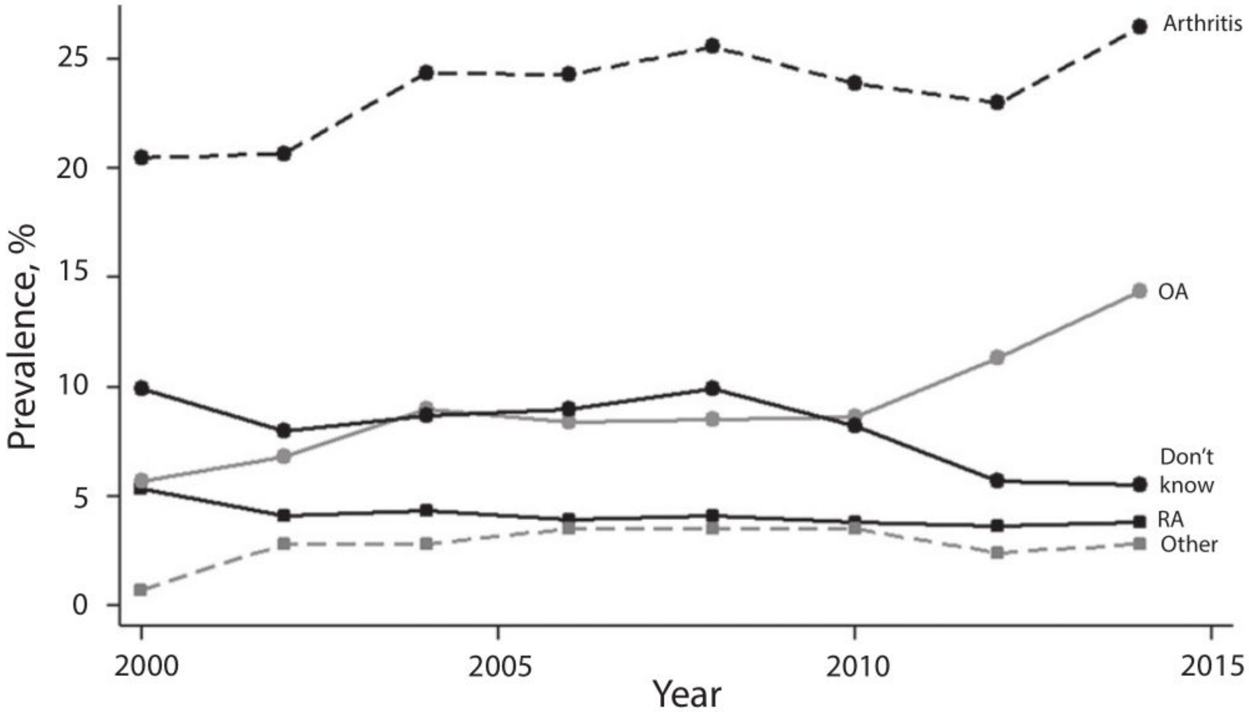
RA-MI example

Various types of arthritis in the united states: prevalence and age-related trends from 1999 to 2014

J Park, A Mendy, ER Vieira - American journal of public ..., 2018 - ajph.aphapublications.org

Objectives. To determine the prevalence trends of osteoarthritis (OA), rheumatoid arthritis (RA), and other types of arthritis in the United States from 1999 to 2014. Methods. We analyzed data on 43 706 community-dwelling adults aged 20 years and older who participated in the 1999–2014 National Health and Nutrition Examination Surveys. We accounted for survey design and sampling weights so that estimates were nationally representative. We assessed temporal trends in age-standardized arthritis prevalence by ...

Cited by 5 Related articles All 11 versions Web of Science: 4

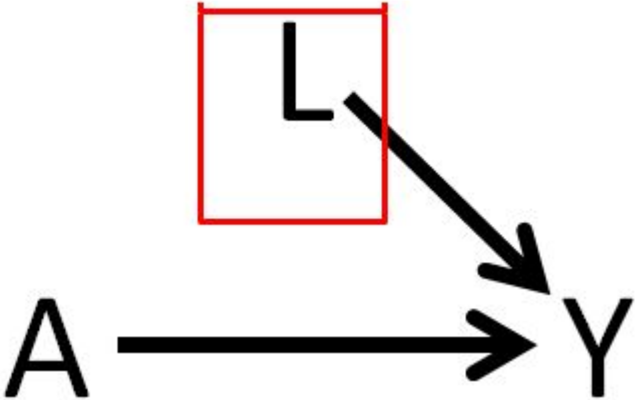


relatively low prevalence of RA; also chronic,

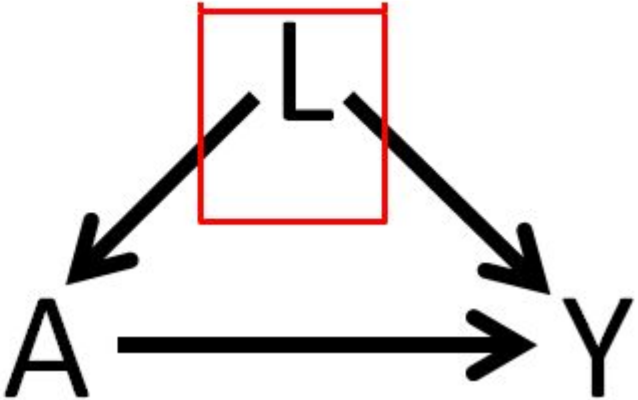
harder to do RCT

RCT vs. Observational study

RCT



Observational studies



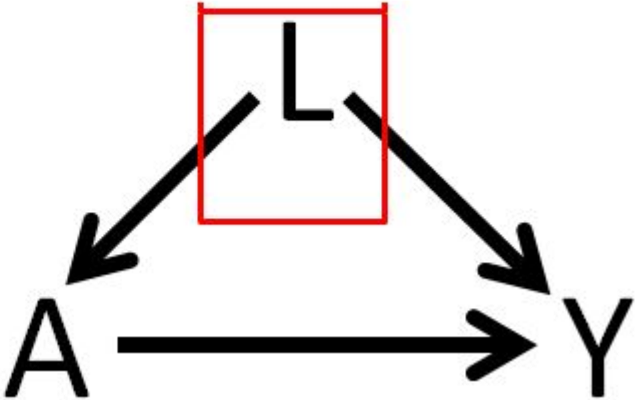
L is not randomized anymore

Notations: confounder

A: Exposure status (1 = exposed; 0 = not)

Y: Outcome

L: Covariates



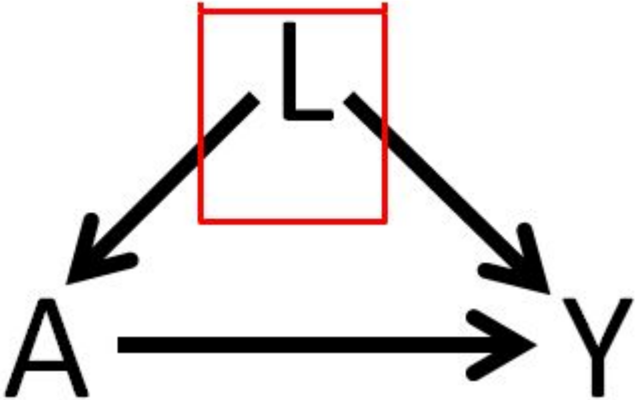
L could be restricted / matched / stratified / adjusted in regression to get unbiased treatment effect

RCT vs. Observational study

Is age really a confounder?

If age distribution in people with rheumatoid arthritis versus people without arthritis is the same, then age is not a confounder (loosely speaking).

Observational studies



RCT vs. Observational study

Is age really a confounder?

age	Non-arthritis	Rheumatoid arthritis
(0,50]	0.63	0.23
(50,70]	0.26	0.52
70+	0.11	0.25

(column heading: arthritis.type)

If age distribution in people with rheumatoid arthritis versus people without arthritis is the same, then age is not a confounder (loosely speaking).

Is that the case here?

Is the age distribution balanced in 2 groups?

Imbalance measure: important concept!! 🏆

- Balance checking is often revealing of variables that require adjustment (responsible for imbalance).

For a continuous variable, the standardized mean difference

$$SMD_{continuous} = \frac{\bar{l}_{RA} - \bar{l}_{NoArthritis}}{\sqrt{\frac{s^2_{RA} + s^2_{NoArthritis}}{2}}}$$

For a binary variable, the standardized proportion difference

$$SMD_{binary} = \frac{\hat{p}_{RA} - \hat{p}_{NoArthritis}}{\sqrt{\frac{\hat{p}_{RA} \times (1 - \hat{p}_{RA}) + \hat{p}_{NoArthritis} \times (1 - \hat{p}_{NoArthritis})}{2}}}$$

Imbalance measure:

Table 1

Stratified by RA

```
##                        Stratified by arthritis.type
##                         Non-arthritis Rheumatoid arthritis SMD
##   n                     4089          325
##   gender = Female (%)   1960 (47.9)   194 (59.7)            0.238
##   diabetes = Yes (%)     358 ( 8.8)    87 (26.8)            0.485
##   smoke = Yes (%)       1796 (43.9)   177 (54.5)            0.212
##   age (%)                                                   0.891
##      (0,50]             2577 (63.0)    74 (22.8)
##      (50,70]            1046 (25.6)   169 (52.0)
##      70+                 466 (11.4)    82 (25.2)
```

SMD = measure of distance between two group means/proportions. SMD >.2 means imbalance.

What to do if imbalance exists?

Logistic(Y ~ A):

crude and potentially biased in the observational setting

crude OR = E[Y(1) vs. Y(0)] = 3.54

Illustrative example: Potential Adjustment variables

Confounders and risk factors (L):

age, BMI, diabetes, smoking.

Demographic variables that could be confounders / risk factors (L):

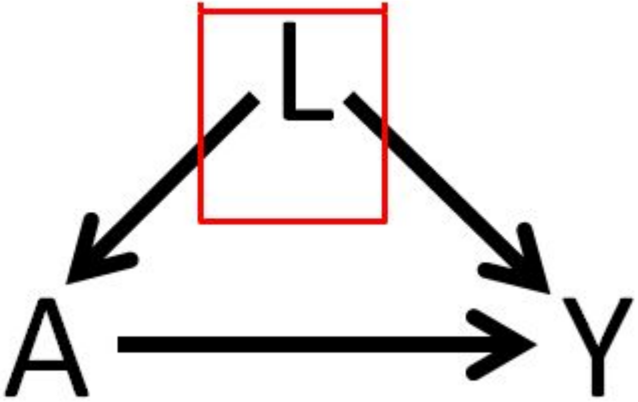
sex, race, education,

marital status, income, origin.

Additional factors / potential confounders (L):

physical activity, access to medical services,

hypertension/high blood pressure and diet



What to do if imbalance exists? Regression

$Logistic(Y \sim A + L1 + L2 + L3 + \ldots Ln)$: adjusted

adjusted OR = E[Y(1) vs. Y(0) | L] = 1.54

Thanks!

ehsan.karim@ubc.ca

www.ehsankarim.com