hmrk 10 - Beimnet Taye

2023-04-09

Worked With Joan Shim and Lucas Yoshida

P1.

1.

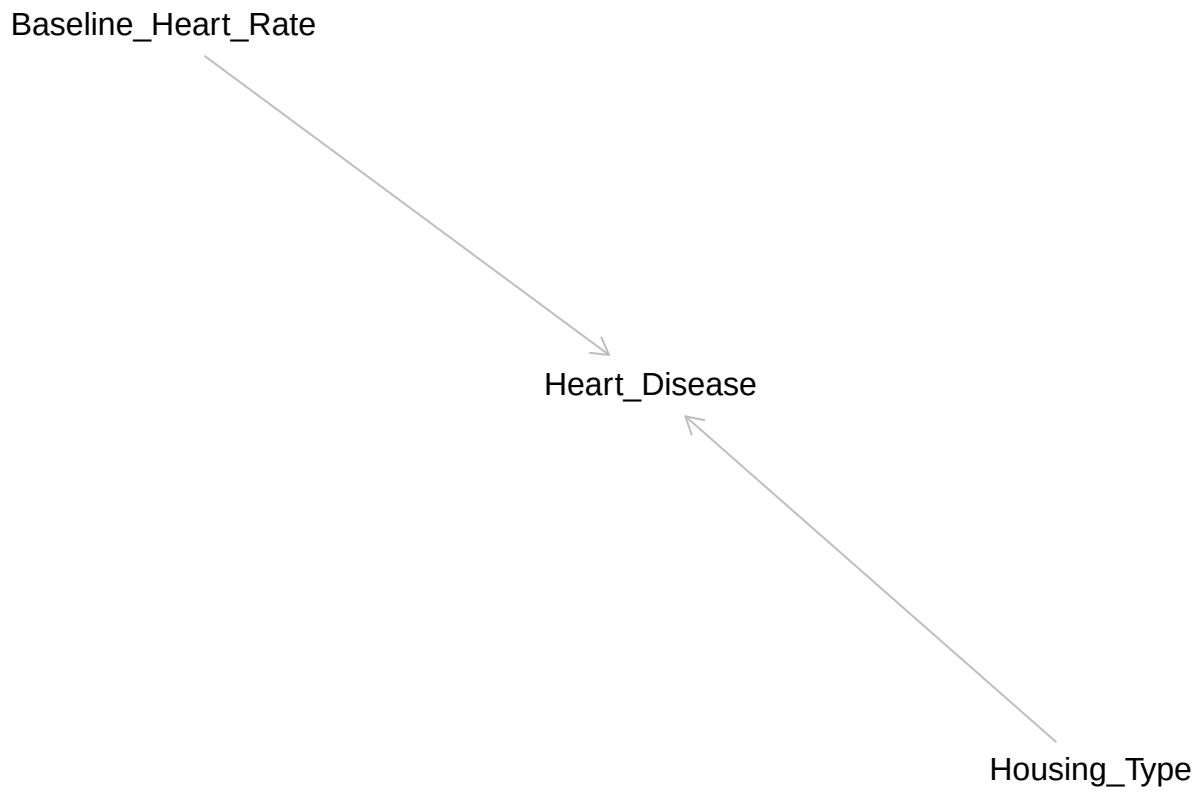
- No they cannot be confounders in this experimental set up since we have randomly assigned the treatment condition. Thus we have probably balanced both known and unknown confounders between the treatment and control groups given a large enough sample size. Without randomization Baseline heart rate would not be a confounder anyways since it doesn't affect our exposure which is corrugated metal use for roofing.

```
dag <- dagitty("dag {


Housing_Type -> Heart_Disease
Baseline_Heart_Rate -> Heart_Disease

               }")
plot(dag)
```

```
## Plot coordinates for graph not supplied! Generating coordinates, see ?coordinates for how to set your
```



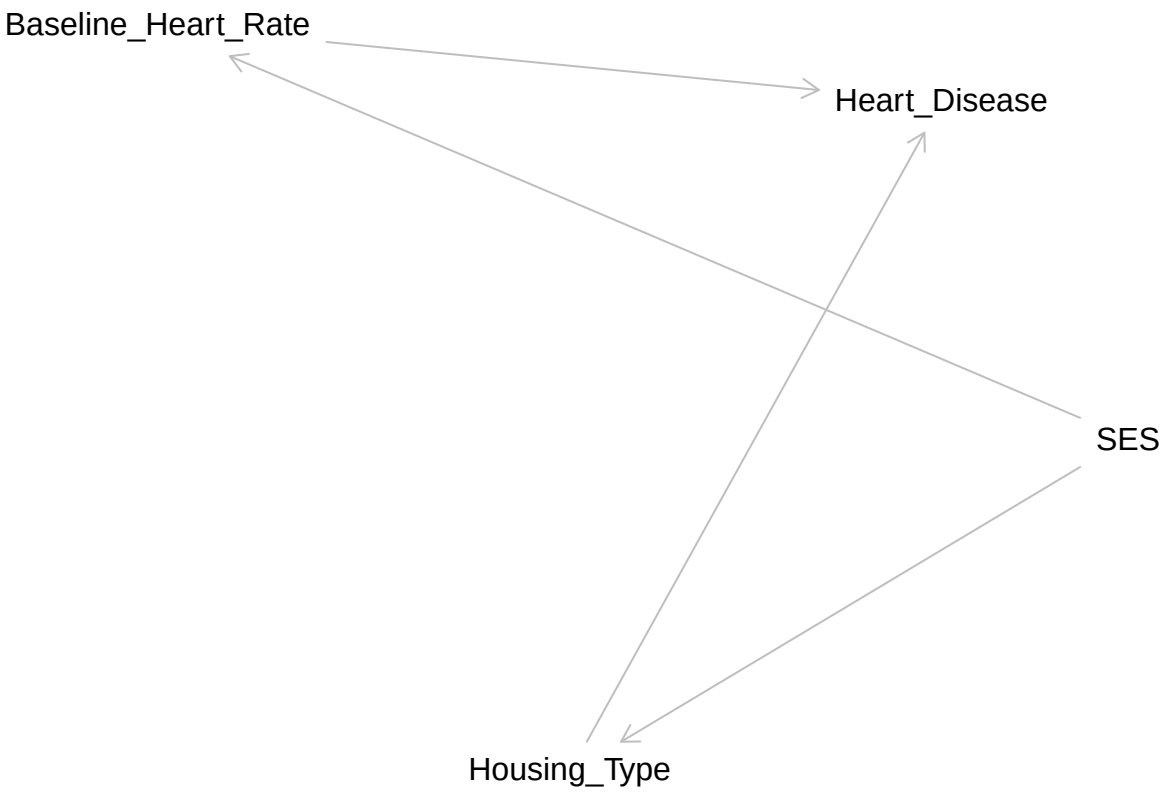
2. - Yes it could be confounder that affects our results since we are not randomizing the housing type anymore and baseline heart rate affects heart disease outcomes and is associated with the exposure of housing type due to having a common cause, SES.

```
dagb <- dagitty("dag {

SES -> Housing_Type
SES -> Baseline_Heart_Rate
Housing_Type -> Heart_Disease
Baseline_Heart_Rate -> Heart_Disease

               }")
plot(dagb)
```

```
## Plot coordinates for graph not supplied! Generating coordinates, see ?coordinates for how to set your
```



3.

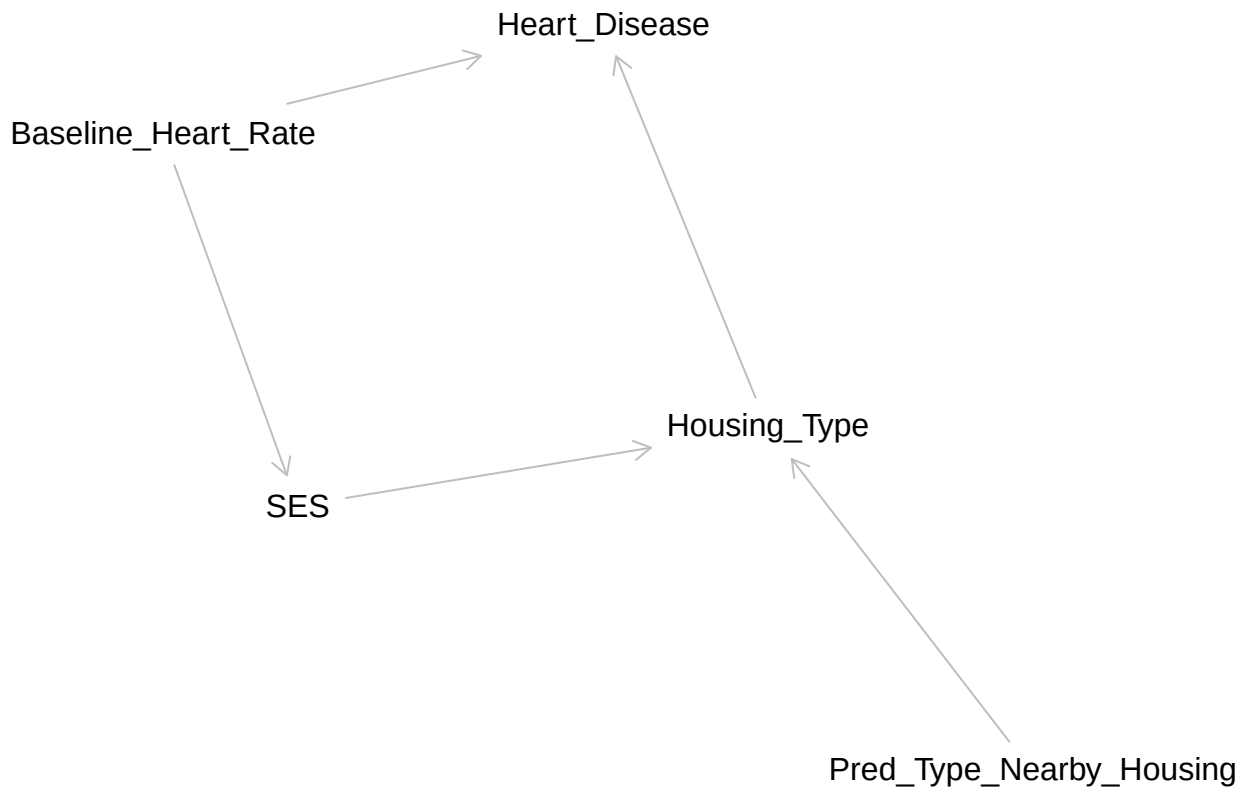
- No since it only affects our exposure and does and affect our outcome since it is in effect randomly assigned.

```
dagc <- dagitty("dag {

Pred_Type_Nearby_Housing -> Housing_Type <- SES <- Baseline_Heart_Rate
Housing_Type -> Heart_Disease <- Baseline_Heart_Rate

               }")
plot(dagc)
```

```
## Plot coordinates for graph not supplied! Generating coordinates, see ?coordinates for how to set you
```



P2

1.

- This is not a good idea since we have a lot of non respondents. Non-respondents could be more similar to each other than to those who did respond resulting in selection bias that would affect our estimate of the true population proportion if we were to only use the data of the respondents. The similarity among non respondents could be due to other variables such as age, sex, zip code, and occupation. For instance if younger people are more likely to use opiates and less likely to respond to the survey than older people who are less likely to use opiates then our estimate would underestimate the true population prevalence.

2.

- At the core of our identification assumption is that responding to the survey (S) is independent of opiate use (Y). This is not the case however if we use the conditional randomization identity conditioning on the other covariates like zip code, sex, and occupation (X). Mathematically:
- For all $P(X, S, Y(1), Y(0))$ where $(Y(s) \perp S|X)$:
- $E[Y(s)] = E[E[Y|X, S = s]]$
- This identification holds true if we have data for every stratum and there are not any unmeasured confounders.

P3.

1.

```
study_data = medicaldata::indo_rct

study_data
```

```
## # A tibble: 602 x 33
##       id site    age  risk gender   outcome sod   pep   recpanc psphinc precut
##    <dbl> <fct> <dbl> <dbl> <fct>    <fct>   <fct> <fct> <fct>   <fct>   <fct>
##  1  1001 1_UM     26   2   1_female 1_yes   1_yes 0_no  1_yes   0_no    0_no
##  2  1002 1_UM     24   1   2_male   0_no    0_no  1_yes 0_no    0_no    0_no
##  3  1003 1_UM     57   1   1_female 0_no    1_yes 0_no  0_no    0_no    0_no
##  4  1004 1_UM     29   2   1_female 1_yes   1_yes 0_no  0_no    0_no    0_no
##  5  1005 1_UM     38   3.5 1_female 0_no    1_yes 1_yes 0_no    1_yes   0_no
##  6  1006 1_UM     59   3   1_female 0_no    1_yes 0_no  0_no    0_no    1_yes
##  7  1007 1_UM     60   1.5 1_female 0_no    0_no  0_no  1_yes   0_no    0_no
##  8  1008 1_UM     29   1   2_male   0_no    0_no  0_no  0_no    0_no    0_no
##  9  1009 1_UM     53   2   2_male   0_no    0_no  0_no  1_yes   0_no    0_no
## 10  1010 1_UM     20   2   2_male   0_no    0_no  0_no  0_no    0_no    1_yes
## # ... with 592 more rows, and 22 more variables: difcan <fct>, pneudil <fct>,
## #   amp <fct>, paninj <fct>, acinar <fct>, brush <fct>, asa81 <fct>,
## #   asa325 <fct>, asa <fct>, prophystent <fct>, therastent <fct>,
## #   pdstent <fct>, sodsom <fct>, bsphinc <fct>, bstent <fct>, chole <fct>,
## #   pbmal <fct>, train <fct>, status <fct>, type <fct>, rx <fct>, bleed <dbl>
```

```
RR <- study_data %>%
  mutate(outcome_num = case_when(
    outcome == "1_yes" ~ 1,
    TRUE ~ 0
        )
  ) %>%
  group_by(rx) %>%
  summarize(risk = mean(outcome_num)) %$% {
    risk[rx == "1_indomethacin"] / risk[rx == "0_placebo"]
  }

RR
```

```
## [1] 0.540352
```

2.

```
plug <- function(mu1,mu0,pi1,pi0) {
  sqrt(
    (mu1/mu0)^2 * (((1-mu0)/(mu0*pi0))+((1-mu1)/(mu1*pi1)))
  )
}
```

```
risks <- study_data %>%
  mutate(outcome_num = case_when(
    outcome == "1_yes" ~ 1,
    TRUE ~ 0
        )
  ) %>%
  group_by(rx) %>%
  summarize(risk = mean(outcome_num))

risks
```

```
## # A tibble: 2 x 2
##   rx               risk
##   <fct>           <dbl>
## 1 0_placebo      0.169
## 2 1_indomethacin 0.0915
```

```
mu0 <- risks %$% {risk[rx == "0_placebo"]}
mu1 <- risks %$% {risk[rx == "1_indomethacin"]}

prob <- study_data %>%
  count(rx) %>%
  mutate(probs = n/sum(n))

p1 <- prob %$% {probs[rx == "1_indomethacin"]}
p0 <- prob %$% {probs[rx == "0_placebo"]}

asymp<- plug(mu1,mu0,p1,p0)
asymp
```

```
## [1] 2.953291
```

```
STE <- asymp/sqrt(nrow(study_data))
STE
```

```
## [1] 0.1203672
```

3.

```
CI <- function(est, asym,data,level = 1.96){
 STE <- asym/sqrt(nrow(data))
  tibble(upper = est + level * STE,
         lower = est - level * STE,
         estimate = est)
}
CI(RR, asymp, data = study_data)
```

```
## # A tibble: 1 x 3
##   upper lower estimate
##   <dbl> <dbl>    <dbl>
## 1 0.776 0.304    0.540
```

4.

- No. We can say that the observable causal estimand that this naive risk ratio is estimating is equal to the causal RR since exposure was randomized thus making treatment (A) independent of outcome (Y). In other words we can use the randomization identity as follows linking an observed ATE estimand with the causal ATE estimand:
- $A \perp Y(1)$
- $E[Y(1)] = E[Y|A = 1]$
- Our calculated naive RR can still be wrong when estimating the observed RR estimand.

5.

- The confidence interval does not include the null of 1 thus the intervention (indomethacin) in this study affects incidence of pancreatits by decreasing it. As a result more testing is needed to see if this drug should be given to the population under the assumption that it reduces incidence of pancreatits but this is a promising start.