Intubation TMLE

Tarragona Datathon

2022-11-11

```r
setwd(dirname(rstudioapi::getSourceEditorContext()$path))

library(tidyverse)
```

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.1 --

## v ggplot2 3.3.5     v purrr   0.3.4
## v tibble  3.1.6     v dplyr   1.0.8
## v tidyr   1.2.0     v stringr 1.4.1
## v readr   2.1.2     v forcats 0.5.1

## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```r
library(lubridate)
```

```
##
## Attaching package: 'lubridate'

## The following objects are masked from 'package:base':
##
##     date, intersect, setdiff, union
```

```r
library(dplyr)
library(tidyr)
library(ggplot2)
# library(dagitty)
# library(ggdag)
library(data.table)
```

```
##
## Attaching package: 'data.table'

## The following objects are masked from 'package:lubridate':
##
##     hour, isoweek, mday, minute, month, quarter, second, wday, week,
##     yday, year

## The following objects are masked from 'package:dplyr':
##
##     between, first, last

## The following object is masked from 'package:purrr':
##
##     transpose
```

```
# install.packages("remotes")
#install.packages("lifecycle")
#install.packages("Rsolnp")
#install.packages("speedglm")
# remotes::install_github("tlverse/tmle3")
# remotes::install_github("tlverse/tmle3mediate")
library(tmle3mediate)
```

```
## tmle3mediate v0.0.3: Targeted Learning for Causal Mediation Analysis
```

```
library(tmle3)
library(sl3)
library(speedglm)
```

```
## Loading required package: Matrix

##
## Attaching package: 'Matrix'

## The following objects are masked from 'package:tidyr':
##
##     expand, pack, unpack

## Loading required package: MASS

##
## Attaching package: 'MASS'

## The following object is masked from 'package:dplyr':
##
##     select
```

```
library(Rsolnp)
library(lifecycle)
```

Import Cleaned Data

```
#tinytex::install_tinytex()

# Load data
cleaned_data <- read_csv("Tab1 H3.csv")
```

```
## New names:
## Rows: 3815 Columns: 18
## -- Column specification
## ---------------------------------------------------------- Delimiter: "," chr
## (2): patientsex, hospital_outcome dbl (16): ...1, X, hospital_coded,
## a_patientid, age, height, weight, bmi, ho...
## i Use `spec()` to retrieve the full column specification for this data. i
## Specify the column types or set `show_col_types = FALSE` to quiet this message.
## * `` -> `...1`
```

Create Variables

Note processing and basic outcomes conducted in Python

```r
cleaned_data$sex_male <- as.numeric(cleaned_data$patientsex == 'M')
cleaned_data$outcome <- as.numeric(cleaned_data$hospital_outcome == "EXITUS")

# Create working dataset
ObsData <-subset(cleaned_data, select=c(age,sex_male,bmi,
                                        sofa.max,sofa.avg,
                                        outcome
                                        ))
ObsData <- na.omit(ObsData)
sapply(ObsData, class)
```

```
##       age  sex_male       bmi  sofa.max  sofa.avg   outcome
## "numeric" "numeric" "numeric" "numeric" "numeric" "numeric"
```

```r
PatsIDs <-subset(cleaned_data, select=c(age,a_patientid,
                                        sex_male,bmi,
                                        sofa.max,sofa.avg,
                                        outcome
                                        ))
PatsIDs <- na.omit(PatsIDs)

head(ObsData,n=2)
```

```
## # A tibble: 2 x 6
##     age sex_male   bmi sofa.max sofa.avg outcome
##   <dbl>    <dbl> <dbl>    <dbl>    <dbl>   <dbl>
## 1    40        1    37        8     4.25       0
## 2    80        0    31        2     1.5        0
```

The Logic of Early Warning Scores

Early warning scores use patient observations to trigger a review by clinicians if they reach a certain threshold Appropriate triggering will allow for early review before significant deterioration, allowing for starting of relevant treatment, and hopefully better outcomes

```r
# dagify(
#   Obs ~ Patient,
#   Trigger ~ Obs,
#   Review ~ Trigger,
#   Treatment ~ Review,
#   Improve ~ Treatment,
#   Deteriorate ~ Treatment
# ) %>%
#   ggplot(aes(x = x, y = y, xend = xend, yend = yend)) +
#   geom_dag_point(size=20) +
#   geom_dag_edges(linemitre=2) +
#   geom_dag_text(size=3)+
#   theme_dag()
#
```

Failure to Trigger Review

Differences in Pulse oximter performance has been shown between ethnicities, due to failure in infrared technology to be calibrated apropriately to differnt skin tones

This device is used in Early Warning Scores to and therefore could result in delayed review and treatment This could result in worse patient outcomes in these groups

```
# dagify(
#   Trigger ~ Obs,
#   Outcome ~ Trigger,
#   Obs ~ Oximeter,
#   Trigger ~ Oximeter
# ) %>%
#   ggplot(aes(x = x, y = y, xend = xend, yend = yend)) +
#   geom_dag_point(size=20) +
#   geom_dag_edges(linemitre=2) +
#   geom_dag_text(size=3)+
#   theme_dag()
```

We therefore investigated the effects of pulse oximetry variable performance on different ethncities Relating this to the TMLE structure this looks like…

```
# dagify(
#   Patient ~ Age + Sex + PMH + Ethnicity,
#   Obs ~ Patient,
#   Outcome ~ Trigger,
#   Trigger ~ Obs,
#   Oximeter ~ Ethnicity,
#   Trigger ~ Obs,
#   Trigger ~ Oximeter
# ) %>%
#   ggplot(aes(x = x, y = y, xend = xend, yend = yend)) +
#   geom_dag_point(size=20) +
#   geom_dag_edges(linemitre=2) +
#   geom_dag_text(size=3)+
#   theme_dag()
```

Prepare the variables

Y: Primary outcome event (1 outcome, 0 no outcome), this includes any of the following: ALIVE EXITUS

A: ("Treatment") -> SEX

Z: The mediating factor (patients' SOFA max score)

(W: comorbidilities) W1: Age (Currently floating number) Could change to categorical ie over under 65

W2: BMI

```
# Outcome
Y  <- "outcome"

# Treatment
A  <- "sex_male"

# Mediators
Z = "sofa.max"

# Covariates
W= c("bmi","age")
```

```r
node_list <- list(
  W = W,
  A = A,
  Z = Z,
  Y = Y
)
```

```r
summary(glm(outcome ~ sex_male + sofa.max + bmi + age,ObsData, family=gaussian(link="identity")))
```

```
##
## Call:
## glm(formula = outcome ~ sex_male + sofa.max + bmi + age, family = gaussian(link = "identity"),
##     data = ObsData)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -0.9082  -0.2347  -0.0656   0.1955   1.0953
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept) -0.1175388  0.0527100  -2.230   0.0259 *
## sex_male     0.0781245  0.0194522   4.016 6.21e-05 ***
## sofa.max     0.0546921  0.0024306  22.501  < 2e-16 ***
## bmi         -0.0093742  0.0015696  -5.972 2.93e-09 ***
## age          0.0059470  0.0005779  10.291  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for gaussian family taken to be 0.1297467)
##
##     Null deviance: 286.25  on 1465  degrees of freedom
## Residual deviance: 189.56  on 1461  degrees of freedom
## AIC: 1173.5
##
## Number of Fisher Scoring iterations: 2
```

Ensemble Learner

Construct an ensemble learner using a handful of popular machine learning algorithms

```r
# SL learners used for continuous data (the nuisance parameter Z)
enet_contin_learner <- Lrnr_glmnet$new(
  alpha = 0.5, family = "gaussian", nfolds = 3
)
lasso_contin_learner <- Lrnr_glmnet$new(
  alpha = 1, family = "gaussian", nfolds = 3
)
fglm_contin_learner <- Lrnr_glm_fast$new(family = gaussian())
mean_learner <- Lrnr_mean$new()
contin_learner_lib <- Stack$new(
  enet_contin_learner, lasso_contin_learner, fglm_contin_learner, mean_learner
)
sl_contin_learner <- Lrnr_sl$new(learners = contin_learner_lib)
```

```
# SL learners used for binary data (nuisance parameters G and E in this case)
enet_binary_learner <- Lrnr_glmnet$new(
  alpha = 0.5, family = "binomial", nfolds = 3
)
lasso_binary_learner <- Lrnr_glmnet$new(
  alpha = 1, family = "binomial", nfolds = 3
)
fglm_binary_learner <- Lrnr_glm_fast$new(family = binomial())
binary_learner_lib <- Stack$new(
  enet_binary_learner, lasso_binary_learner, fglm_binary_learner, mean_learner
)
sl_binary_learner <- Lrnr_sl$new(learners = binary_learner_lib)

# create list for treatment and outcome mechanism regressions
learner_list <- list(
  Y = sl_contin_learner,
  A = sl_binary_learner
)
```

Targeted Estimation of the Natural Indirect Effect

```
require(methods)

tmle_spec_NIE <- tmle_NIE(
  e_learners = Lrnr_cv$new(lasso_binary_learner, full_fit = TRUE),
  psi_Z_learners = Lrnr_cv$new(lasso_contin_learner, full_fit = TRUE),
  max_iter = 1
)
ObsData_NIE <- tmle3(
  tmle_spec_NIE, ObsData, node_list, learner_list
)
ObsData_NIE
```

```
## A tmle3_Fit that took 1 step(s)
##    type                  param  init_est   tmle_est         se       lower
## 1:  NIE NIE[Y_{A=1} - Y_{A=0}] 0.0175191 0.01716855 0.01376932 -0.00981883
##         upper psi_transformed lower_transformed upper_transformed
## 1: 0.04415592      0.01716855       -0.00981883        0.04415592
```

Based on the output, we see that the indirect effect of the treatment through the mediators (sex male) is 0.008197 - IMV cohort Based on the output, we see that the indirect effect of the treatment through the mediators (sex male) in all patients H3 is 0.01737

Targeted Estimation of the Natural Direct Effect

```
tmle_spec_NDE <- tmle_NDE(
  e_learners = Lrnr_cv$new(lasso_binary_learner, full_fit = TRUE),
  psi_Z_learners = Lrnr_cv$new(lasso_contin_learner, full_fit = TRUE),
  max_iter = 1
)
ObsData_NDE <- tmle3(
  tmle_spec_NDE, ObsData, node_list, learner_list
```

```
)
ObsData_NDE
```

```
## A tmle3_Fit that took 1 step(s)
##    type                  param   init_est   tmle_est         se      lower
## 1:  NDE NDE[Y_{A=1} - Y_{A=0}] 0.06587429 0.06588656 0.01513195 0.03622847
##         upper psi_transformed lower_transformed upper_transformed
## 1: 0.09554465      0.06588656        0.03622847        0.09554465
```

From this, we can draw the conclusion that the direct effect of the treatment (through all paths not involving the mediators (ethnicity)) is 0.09847

Together, the estimates of the natural direct and indirect effects approximately recover the average treatment effect, that is, based on these estimates of the NDE and NIE, the ATE is roughly .

References

https://tlverse.org/tlverse-handbook/causal-mediation-analysis.html https://migariane.github.io/TMLE.nb.html