preprocessing

Jennifer Ci, Thu Vu, Lily Hanyi Wang

2022-10-26

Dataset `TEVAR_PROC_07` to start with.

The population of interest in the asymptomatic and symptomatics groups.

`PRESENTATION`: $0 =$ Asymptomatic,$1 =$ Symptomatic,$2 =$ Rupture

```
## Get nicer `table1` LaTeX output by simply installing the `kableExtra` package
```

	Overall
	(N=19564)
PRESENTATION	
Asymptomatic	9272 (47.4%)
Symptomatic	6624 (33.9%)
Rupture	1168 (6.0%)
Missing	2500 (12.8%)

table for some potential outcome variables

```
## Get nicer `table1` LaTeX output by simply installing the `kableExtra` package
```

	Asymptomatic	Symptomatic	Rupture	Overall
	(N=9272)	(N=6624)	(N=1168)	(N=19564)
factor(DEAD)				
0	8103 (87.4%)	5570 (84.1%)	741 (63.4%)	16254 (83.1%)
1	1168 (12.6%)	1053 (15.9%)	427 (36.6%)	3295 (16.8%)
Missing	1 (0.0%)	1 (0.0%)	0 (0%)	15 (0.1%)
PROC_SURVIVALDAYS				
Mean (SD)	719 (725)	657 (730)	522 (709)	798 (883)
Median [Min, Max]	456 [-355, 3360]	407 [0, 3200]	215 [0, 3410]	454 [-355, 3970]
Missing	1 (0.0%)	0 (0%)	0 (0%)	1 (0.0%)

table for some other potential variables of interest

```
## Get nicer `table1` LaTeX output by simply installing the `kableExtra` package
```

	Asymptomatic	Symptomatic	Rupture	Overall
	(N=9272)	(N=6624)	(N=1168)	(N=19564)
AGE				
Mean (SD)	70.3 (11.5)	62.3 (16.1)	64.7 (19.5)	66.8 (14.6)
Median [Min, Max]	72.0 [0, 90.0]	65.0 [0, 90.0]	71.0 [0, 90.0]	70.0 [0, 90.0]
factor(PREOP_SMOKING)				

	Asymptomatic	Symptomatic	Rupture	Overall
0	1795 (19.4%)	2266 (34.2%)	458 (39.2%)	5132 (26.2%)
1	4818 (52.0%)	1975 (29.8%)	330 (28.3%)	8148 (41.6%)
2	2653 (28.6%)	2350 (35.5%)	340 (29.1%)	6145 (31.4%)
Missing	6 (0.1%)	33 (0.5%)	40 (3.4%)	139 (0.7%)

Code Appendix

```r
knitr::opts_chunk$set(echo = FALSE)
library(tidyverse)
library(table1)


## ------------- working directories for Lily ----------
wd_lily = '/Users/hanyiwang/Desktop/Comparative-analysis-of-treatments-of-CAA'
path_lily = c(
  "../data/TEVAR_International_20210712/TEVAR_International_LTF_r12_2_14_20210701.csv",
  "../data/TEVAR_International_20210712/TEVAR_International_PROC_r12_2_14_20210701.csv",
  "../data/TEVAR_International_20210901/TEVAR_International_LTF_r12_2_14_20210901.csv",
  "../data/TEVAR_International_20210901/TEVAR_International_PROC_r12_2_14_20210901.csv")


## ------------- read data ----------
setwd(wd_lily)
TEVAR_LTF_07 = read.csv(path_lily[1])
TEVAR_PROC_07 = read.csv(path_lily[2])
#TEVAR_LTF_09 = read.csv(path_lily[3])
#TEVAR_PROC_09 = read.csv(path_lily[4])
## ------------- data cleaning----------
TEVAR_PROC_07 = TEVAR_PROC_07 %>%
  mutate(PRESENTATION = factor(PRESENTATION,levels = c(0,1,2),
                               labels = c('Asymptomatic','Symptomatic','Rupture')))

## ------------- population of interest ----------
table1(~  PRESENTATION, data = TEVAR_PROC_07)
## ------------- descriptive statistics table for outcomes----------
table1(~ factor(DEAD) + PROC_SURVIVALDAYS | PRESENTATION, data = TEVAR_PROC_07)
## ------------- descriptive statistics table for variables of interest----------
table1(~ AGE+factor(PREOP_SMOKING) | PRESENTATION, data = TEVAR_PROC_07)
```