preprocessing

Jennifer Ci, Thu Vu, Lily Hanyi Wang

2022-10-26

Dataset `TEVAR_PROC_07` to start with.

population of interest: the asymptomatic and symptomatics groups.

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

What does rupture mean?

Should we just delete data missing symptom info?

`PATHOLOGY`: 1 = Aneurysm,2 = Dissection,3 = Aneurysm from dissection,4 = Trauma,5 = Penetrating Ulcer (PAU),6 = Intramural Hematoma (IMH),7 = PAU with IMH,8 = Aortic Thrombus,9 = Other (Retired) (retired since 09/30/2014),10 = Aorto-esophageal Fistula (Retired) (retired since 09/30/2014),11 = Aorto-bronchial Fistula (Retired) (retired since 09/30/2014)

Old variable, used to be called Indication, now mapped to Pathology. Mapping detail: 1 = TAA->1 = Aneurysm; 2 = TAAA->1 = Aneurysm; 4 = Dissection->2 = Dissection; 3=Aneurysm from dissection; 3 = Trauma->4=Trauma; 5 = Penetrating Ulcer->5 = Penetrating Ulcer (PAU); 6 = Aortic Intramural Hematoma->6 = Intramural Hematoma (IMH); 7=PAU with IMH; 8=Aortic Thrombus

Type of AA?

	Overall
	(N=19564)
PATHOLOGY	
1	10351 (52.9%)
2	4194 (21.4%)
3	1048 (5.4%)
4	1528 (7.8%)
5	1051 (5.4%)
6	397 (2.0%)
7	363 (1.9%)
8	133 (0.7%)
9	375 (1.9%)
10	3 (0.0%)
11	1 (0.0%)
Missing	120 (0.6%)
PRESENTATION	
Asymptomatic	9272 (47.4%)
Symptomatic	6624 (33.9%)
Rupture	1168 (6.0%)
Missing	2500 (12.8%)

potential outcome variables

`PROC_SURVIVALDAYS`: This should be the longest known time of survival data available for the patient. Survival days are calculated as the Last Date of Contact (or Date of Death) for the patient - Procedure date for a procedure. Please refer to included Death and Survival Days Logic.pdf for additional details.”

survival analysis

`POSTOP_LOS`: Length of Stay in days calculated by DISCHARGE_DT - SURGERY_DT

endoleak what does that mean

	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%)
POSTOP_LOS				
Mean (SD)	6.48 (28.4)	10.0 (20.2)	16.0 (51.1)	8.57 (28.0)
Median [Min, Max]	3.00 [0, 1100]	6.00 [0, 1100]	9.00 [0, 1140]	5.00 [0, 1140]
Missing	1 (0.0%)	1 (0.0%)	0 (0%)	4 (0.0%)

patient condition variables:

`AGECAT`: 1 = <40,2 = 40-49,3 = 50-59,4 = 60-69,5 = 70-79,6 = 80-89,7 = >89

`GENDER`: 1 = Male,2 = Female

`PRIOR_CVD`: 0 = None,1 = hx stroke, asymptomatic,2 = hx stroke, minor deficit,3 = hx stroke, major deficit

`PRIOR_CAD`: 0 = None,1 = hx MI but no sx,2 = Stable angina,3 = Unstable angina or MI < 6 mos (retired since 09/12/2012),4 = MI < 6 mos,5 = Unstable angina

`PRIOR_CHF`: 0 = None,1 = Asymp, hx CHF,2 = Mild,3 = Moderate,4 = Severe

`COPD`: 0 = No,1 = Not Treated,2 = On Meds,3 = On Home Oxygen

`DIABETES`: 0 = None,1 = Diet,2 = Non-insulin Meds,3 = Insulin

`HTN`: History of hypertension; 0 = No, 1 = Yes (>=140/90 or history) (retired since 11/15/2016), 2 = Yes, controlled [added on 04/13/2020] , 3 = Yes, uncontrolled [added on 04/13/2020]

`PREOP_SMOKING`: 0 = Never,1 = Prior,2 = Current

`PRIOR_AORSURG`: Any aortic procedures performed on a separate date prior to the index procedure; 0 = None,1 = Open,2 = Endo,3 = Both,4 = Other (retired since 09/30/2014)

	Asymptomatic	Symptomatic	Rupture	Overall
	(N=9272)	(N=6624)	(N=1168)	(N=19564)
GENDER				
male	6356 (68.6%)	3988 (60.2%)	736 (63.0%)	12667 (64.7%)
female	2916 (31.4%)	2636 (39.8%)	432 (37.0%)	6897 (35.3%)
AGE				
Mean (SD)	70.3 (11.5)	62.3 (16.1)	64.7 (19.5)	66.8 (14.6)

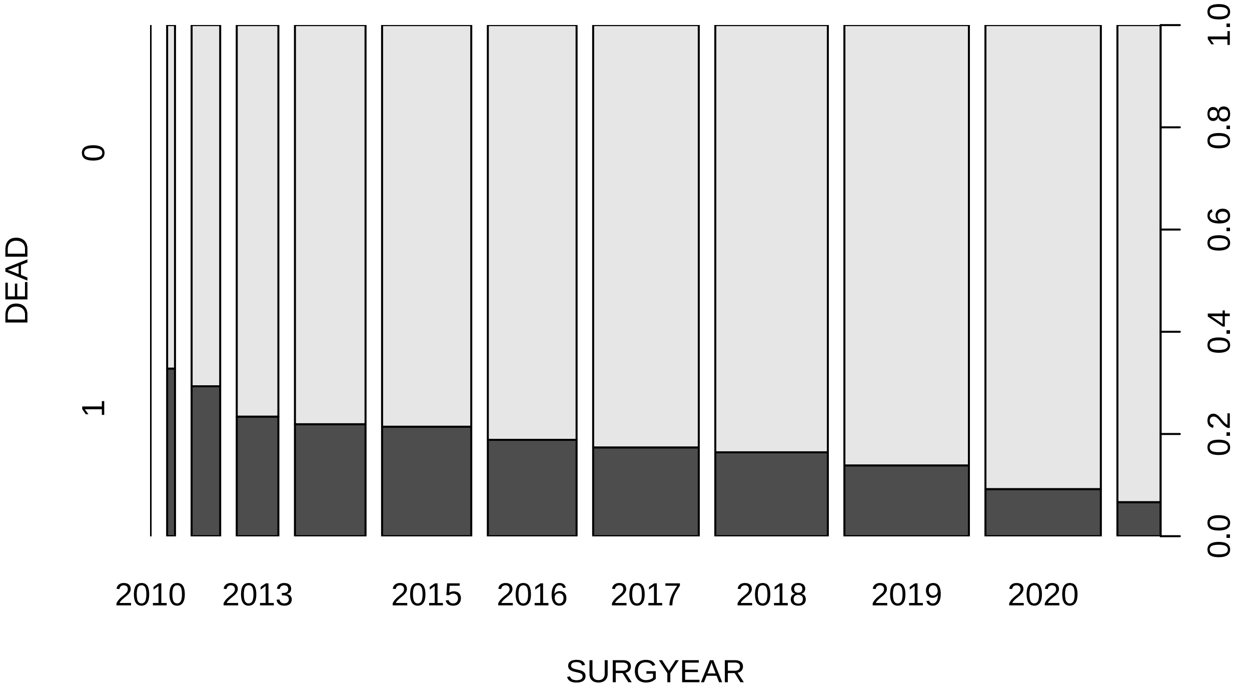
	Asymptomatic	Symptomatic	Rupture	Overall
Median [Min, Max]	72.0 [0, 90.0]	65.0 [0, 90.0]	71.0 [0, 90.0]	70.0 [0, 90.0]
AGECAT				
<40	221 (2.4%)	662 (10.0%)	170 (14.6%)	1198 (6.1%)
40-49	252 (2.7%)	698 (10.5%)	71 (6.1%)	1185 (6.1%)
50-59	809 (8.7%)	1181 (17.8%)	123 (10.5%)	2411 (12.3%)
60-69	2423 (26.1%)	1633 (24.7%)	197 (16.9%)	4826 (24.7%)
70-79	3795 (40.9%)	1552 (23.4%)	323 (27.7%)	6543 (33.4%)
80-89	1663 (17.9%)	814 (12.3%)	241 (20.6%)	3138 (16.0%)
>89	109 (1.2%)	84 (1.3%)	43 (3.7%)	263 (1.3%)
factor(PREOP_SMOKING)				
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%)
factor(PRIOR_CVD)				
0	8196 (88.4%)	5949 (89.8%)	1044 (89.4%)	15348 (78.5%)
1	661 (7.1%)	401 (6.1%)	65 (5.6%)	1135 (5.8%)
2	329 (3.5%)	197 (3.0%)	24 (2.1%)	551 (2.8%)
3	76 (0.8%)	55 (0.8%)	18 (1.5%)	150 (0.8%)
Missing	10 (0.1%)	22 (0.3%)	17 (1.5%)	2380 (12.2%)
factor(PRIOR_CAD)				
0	7072 (76.3%)	5571 (84.1%)	991 (84.8%)	15560 (79.5%)
1	1627 (17.5%)	677 (10.2%)	111 (9.5%)	2777 (14.2%)
2	454 (4.9%)	213 (3.2%)	30 (2.6%)	798 (4.1%)
3	0 (0%)	0 (0%)	0 (0%)	7 (0.0%)
4	66 (0.7%)	68 (1.0%)	14 (1.2%)	168 (0.9%)
5	44 (0.5%)	73 (1.1%)	10 (0.9%)	153 (0.8%)
Missing	9 (0.1%)	22 (0.3%)	12 (1.0%)	101 (0.5%)
factor(PRIOR_CHF)				
0	7982 (86.1%)	5885 (88.8%)	1009 (86.4%)	17044 (87.1%)
1	786 (8.5%)	414 (6.3%)	78 (6.7%)	1432 (7.3%)
2	297 (3.2%)	166 (2.5%)	37 (3.2%)	574 (2.9%)
3	172 (1.9%)	103 (1.6%)	23 (2.0%)	329 (1.7%)
4	30 (0.3%)	36 (0.5%)	6 (0.5%)	90 (0.5%)
Missing	5 (0.1%)	20 (0.3%)	15 (1.3%)	95 (0.5%)
factor(COPD)				
0	6128 (66.1%)	5141 (77.6%)	898 (76.9%)	13976 (71.4%)
1	797 (8.6%)	432 (6.5%)	75 (6.4%)	1487 (7.6%)
2	1867 (20.1%)	837 (12.6%)	143 (12.2%)	3211 (16.4%)
3	474 (5.1%)	195 (2.9%)	41 (3.5%)	801 (4.1%)
Missing	6 (0.1%)	19 (0.3%)	11 (0.9%)	89 (0.5%)
factor(DIABETES)				
0	7698 (83.0%)	5614 (84.8%)	978 (83.7%)	16316 (83.4%)
1	371 (4.0%)	267 (4.0%)	53 (4.5%)	790 (4.0%)
2	921 (9.9%)	491 (7.4%)	75 (6.4%)	1698 (8.7%)
3	277 (3.0%)	233 (3.5%)	49 (4.2%)	671 (3.4%)
Missing	5 (0.1%)	19 (0.3%)	13 (1.1%)	89 (0.5%)
factor(HTN)				
0	1158 (12.5%)	1177 (17.8%)	323 (27.7%)	3040 (15.5%)
1	6319 (68.2%)	4107 (62.0%)	634 (54.3%)	13118 (67.1%)
2	1274 (13.7%)	706 (10.7%)	126 (10.8%)	2114 (10.8%)
3	483 (5.2%)	571 (8.6%)	63 (5.4%)	1117 (5.7%)

	Asymptomatic	Symptomatic	Rupture	Overall
Missing	38 (0.4%)	63 (1.0%)	22 (1.9%)	175 (0.9%)
factor(PRIOR_AORSURG)				
0	6753 (72.8%)	5347 (80.7%)	954 (81.7%)	14839 (75.8%)
1	1377 (14.9%)	719 (10.9%)	108 (9.2%)	2549 (13.0%)
2	917 (9.9%)	479 (7.2%)	92 (7.9%)	1654 (8.5%)
3	216 (2.3%)	63 (1.0%)	10 (0.9%)	289 (1.5%)
4	0 (0%)	0 (0%)	1 (0.1%)	131 (0.7%)
Missing	9 (0.1%)	16 (0.2%)	3 (0.3%)	102 (0.5%)

other variables?

would surgery year affect? eg.progression of surgery?

increasing number of surgeries done. decreasing death rate.



Medical center info:

19 regions, 189 centers, 1094 physicians.

Most physicians only performed 1 or 2 procedures. Several performed over 100 procedures? Is that real?

There are regions and centers that perfomed many procedures. Would that affect the outcome?

Code Appendix

```
knitr::opts_chunk$set(echo = FALSE,message = FALSE,warning = 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(DEAD=factor(DEAD)) %>%
  mutate(PRESENTATION = factor(PRESENTATION,levels = c(0,1,2),
                               labels = c('Asymptomatic','Symptomatic','Rupture'))) %>%
  mutate(AGECAT = factor(AGECAT,levels = c(1,2,3,4,5,6,7),
                         labels = c('<40','40-49','50-59','60-69','70-79','80-89','>89'))) %>%
  mutate(GENDER=factor(GENDER,levels=c(1,2),
                       labels=c('male','female'))) %>%
  mutate(SURGYEAR=factor(SURGYEAR)) %>%
  mutate(PATHOLOGY=factor(PATHOLOGY))

## ------------- population of interest ----------
table1(~  PATHOLOGY+PRESENTATION, data = TEVAR_PROC_07)
## ------------- descriptive statistics table for outcomes----------
table1(~ factor(DEAD) + PROC_SURVIVALDAYS+POSTOP_LOS | PRESENTATION, data = TEVAR_PROC_07)
## ------------- descriptive statistics table for patients conditions----------
table1(~ GENDER+AGE+AGECAT+factor(PREOP_SMOKING)+factor(PRIOR_CVD)+factor(PRIOR_CAD)+
         factor(PRIOR_CHF)+factor(COPD)+factor(DIABETES)+factor(HTN)+factor(PRIOR_AORSURG)
       | PRESENTATION, data = TEVAR_PROC_07)


plot(DEAD~SURGYEAR,data=TEVAR_PROC_07)
## ------------- descriptive statistics table for other variables of interest----------

#TEVAR_PROC_07 %>% select(REGIONID) %>% table()
#TEVAR_PROC_07 %>% select(CENTERID) %>% table()
#TEVAR_PROC_07 %>% select(PHYSICIANID) %>% table()
```