# Data Preprocessing for the VQI TEVAR Dataset

### Jennifer Ci, Thu Vu, Lily Hanyi Wang

## Datasets Merging

Compare the data from July 2021 and September 2021. Keep the most updated ones.

There are 6 rows from LTF July data missing in September dataset. And also 6 missing from the PROC dataset.

Variables that exists in both LTF and PROC datasets are: PATIENTID, PRIMPROCID, DEAD, PROC_SURVIVALDAYS. Merge by these variables.

In the long term followup dataset, the variables used in this project are: `PATIENTID`, `DEAD`, `PROC_SURVIVALDAYS`, `LTF_NUM_REINT`

Initially, there are in total 25211 objectives in the dataset.

## Exclusion criteria:

- `PRESENTATION`: exclude rupture patients
- `PATHOLOGY`: exclude groups with pathology: 4 = trauma, 8 = Aortic Thrombus,9 = Other,10 = Aorto-esophageal Fistula,11 = Aorto-bronchial Fistula
- `URGENCY`: exclude rupture. (`elective` is the same as `asymptomatic`)
- `PROXZONE_DISEASE`: exclude 0 and 1
- `DISTZONE_DISEASE`: exclude 0
- `PROXZONE_DISEASE < DISTZONE_DISEASE`: disease starting point should be earlier than ending point. 35 wrong data points with distal zone < proximal zone are excluded.

After exclsion, there are in total 15424 objectives in the final `TEVAR` dataset. 1995 participants have more than one records in this dataset.

## Data Cleaning

### Patient demographic and co-morbidities

- Comorbidity history variables: changed to Yes/No scale.
- `PREOP_CREAT`: merge `PREOP_CREAT` with retired `R_CR_PRESENT` (mg/dL)

### Operative variables

- `PATHOLOGY`: merge levels PAU and IMH
- `URGENCY`: duplicate with `PRESENTATION` and doesn't make sense, but leave it there.
- `extent`: type of TAAA based on certain criteria calculated by `PROXZONE_DISEASE` and `DISTZONE_DISEASE`.
- `ILIACDEV_END`: from merging `ILIACDEV_END_R`, `ILIACDEV_END_L`

- **ACCESS:** from merging `ACCESS_R`, `ACCESS_L`: Percutaneous if both are Percutaneous, Open o.w.
- **DEV_GTYPE:** merge `DEV1_GTYPE`, `DEV2_GTYPE`, `DEV3_GTYPE`: If one device is 'Custom' or 'Physician modified', classified to this instead of 'Standard'

**Filter FBVAR patients based on having at least one branch, re-leveled as following.**

- `lrenal`: re-leveled `BRANCH_LRENAL_TRT`, retired version `R_LT_RENAL` ignored.
- `rrenal`: re-leveled `BRANCH_RRENAL_TRT`, retired version `R_RT_RENAL` ignored.
- `sma`: re-leveled `BRANCH_SMA_TRT`, retired version `R_SMA` ignored.
- `celiac`: re-leveled `BRANCH_CELIAC_TRT`, retired version `R_CELIAC` ignored.
- `lsub`: re-leveled `BRANCH_LSUB_TRT`, retired version `R_L_SUBCLAV` ignored.

Current levels: 0 = None,1 = Purposely covered,2 = Unintentionally covered,3 = Occluded - coil,4 = Occluded - plug,5 = Occluded - open,6 = Stent,7 = Stent-graft,8 = Chimney,9 = Scallop,10 = Stented Scallop,11 = Fenestration,12 = Stented-fen,13 = Fen branch,14 = Side-arm branch,15 = Surgical bypass,16 = Thromboembolectomy,17 = Iliac Device

**ignore some retired variables without current version**

- `R_DISTATTZONE`: Distal Attachment Zone
- `R_GDPROXIMAL`: Graft Diameter Proximal
- `R_GRFTCONFIG`: Graft Configuration
- `R_PRATTZONE`: Prox. Attachment Zone

Among the whole dataset, we filtered out 3295 FBVAR patients for our study.

**Outcomes**

- `POSTOP_LOS`: changed into binary, more than a week or not.
- Create `POSTOP_AH`: Combine `POSTOP_AH`, `POSTOP_MI`,`POSTOP_DYSRHYTHMIA` for post-procedure abnormal heart disease
- Create `BRANCH_POST`: `BRANCH_XXX_POST` changed to Yes/No scale. Then combine `BRANCH_LSUB_POST`, `BRANCH_CELIAC_POST`,`BRANCH_SMA_POST`,`BRANCH_RRENAL_POST`,`BRANCH_LRENAL_POST`,`BRANCH_RCOMILI_POST`, `BRANCH_LCOMILI_POST`

**update some variables with current version**

- `R_ENDOLEAK_AT_COMPLETION` => `LEAKATCOM_XXX` variables. Only use `LEAKATCOMP_NONE`. Ignore others or have a brief look
- `R_POSTOP_HEMATOMA` => `ACCESS_HEMATOMA_R`, `ACCESS_HEMATOMA_L`; `R_POSTOP_SITEOCC` => `ACCESS_OCCLUSION_R`, `ACCESS_OCCLUSION_L`. Merge hematoma and occlusion, create new variable: `ACCESS_COMPLICATION`
- `R_POSTOP_SSI` => `ACCESS_INFECTION_R`, `ACCESS_INFECTION_L` Merge left and right, create new variable: `ACCESS_INFECTION`

**ignore some retired version variables, only use the current version**

- `R_POSTOP_BOWELISCH` <= `POSTOP_INTISCH`: Bowel Ischemia
- `R_LE_ISCH` <= `POSTOP_LEGEMBO`: LE Ischemia
- `R_POSTOP_RENAL` <= `POSTOP_DIALYSIS`: change of renal function

**record treatment status of the vessels**

- `NUM_TREATED_BRANCHES`: number of treated branches: 4,3,2,1

- `NUM_TREATED_RENALS`: number of treated renals: 2,1,0

- `OCCLUDED_RENAL,OCCLUDED_SMA,OCCLUDED_CELIAC,OCCLUDED_LSUB`: whether this vessel is occluded or covered.

**Others**

Variables we cleaned but are not needed after discussion goes here.

**Volume variables**

## Store a new dataset for further study

Select the variable related to our study. Give them labels for better-looking tables. Finally store the new dataset as a seperate csv file, so that we could use the cleaned dataset in the future modeling.

## Code Appendix

```r
knitr::opts_chunk$set(echo = FALSE,message = FALSE,warning = FALSE)
library(tidyverse)
library(Hmisc)
library(table1)
library(ggplot2)
## ------------- 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",
  "../data/FBVAR.csv")


## ------------- working directories for Jenn ----------
wd_jenn = '/Users/jenniferci/Desktop/Comparative-analysis-of-treatments-of-CAA-main'
path_jenn = c(
  "/Users/jenniferci/Desktop/Comparative-analysis-of-treatments-of-CAA-main/TEVAR_International_2021071
  "/Users/jenniferci/Desktop/Comparative-analysis-of-treatments-of-CAA-main/TEVAR_International_2021071
  "/Users/jenniferci/Desktop/Comparative-analysis-of-treatments-of-CAA-main/TEVAR_International_2021090
  "/Users/jenniferci/Desktop/Comparative-analysis-of-treatments-of-CAA-main/TEVAR_International_2021090



## ------------- 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])

#setwd(wd_jenn)
#TEVAR_LTF_07 = read.csv(path_jenn[1])
#TEVAR_PROC_07 = read.csv(path_jenn[2])
#TEVAR_LTF_09 = read.csv(path_jenn[3])
#TEVAR_PROC_09 = read.csv(path_jenn[4])
## ------------- merge July and September data ----------

# find data in LTF July data but not in LTF September data by `PATIENTID`
# add these data points to the September data
TEVAR_LTF <- rbind(TEVAR_LTF_07[! TEVAR_LTF_07$PATIENTID  %in% TEVAR_LTF_09$PATIENTID,],
                   TEVAR_LTF_09)

# filter the varaibles in the LTF dataset
TEVAR_LTF <- TEVAR_LTF %>% select(PATIENTID,PRIMPROCID,DEAD,PROC_SURVIVALDAYS,LTF_NUM_REINT)

# Similar for PROC data
TEVAR_PROC <-rbind(TEVAR_PROC_07[! TEVAR_PROC_07$PATIENTID  %in% TEVAR_PROC_09$PATIENTID,],
                   TEVAR_PROC_09)

## ------------- merge LTF and PROC data----------
# same variables in LTF and PROC data
#colnames(TEVAR_PROC)[colnames(TEVAR_PROC) %in% colnames(TEVAR_LTF)]
```

```r
TEVAR <- merge(TEVAR_LTF,TEVAR_PROC, all = TRUE,
               by=c("PATIENTID","PRIMPROCID","DEAD","PROC_SURVIVALDAYS"))

## ------------- exclusion----------
TEVAR = TEVAR %>%
  filter(PRESENTATION !=2) %>%
  filter(PATHOLOGY %in% c(1,2,3,5,6,7)) %>%
  filter(URGENCY %in% c(1,2,3)) %>%
  filter(PROXZONE_DISEASE %in% c(2,3,4,5,6,7,8,9)) %>%
  filter(DISTZONE_DISEASE > PROXZONE_DISEASE)

n_occur <- data.frame(table(TEVAR$PATIENTID))
#n_occur[n_occur$Freq > 1,]%>%n_distinct

## ------------- variables cleaning, patient demographic and co-morbidities----------
TEVAR = TEVAR %>%
  mutate(AGECAT = factor(AGECAT,levels = c(1,2,3,4,5,6,7),
                         labels = c('<50','<50','50-59','60-69','70-79','>79','>79'))) %>%
  mutate(URGENCY=factor(URGENCY,levels = c(1,2,3),labels = c('Elective','Urgent','Emergent'))) %>%
  mutate(GENDER=factor(GENDER,levels=c(1,2),labels=c('male','female'))) %>%
  mutate(R_PREOP_AMBUL = factor(R_PREOP_AMBUL,levels = c(1,2,3,4),
                                labels=c("Amb","Amb w/ Assistance","Wheelchair","Bedridden"))) %>%
  mutate(ETHNICITY = factor(ETHNICITY,levels=c(0,1),
                            labels = c('None Hispanic or Latino','Hispanic or Latino'))) %>%
  mutate(RACE=factor(RACE,levels = c(5,3,2,1,4,6,7),
                     labels = c('White','Black or African American','Asian',
                                'American Indian or Alaskan Native',
                                'Native Hawaiian or other Pacific Islander','More than 1 race',
                                'Unknown/Other'))) %>%
  mutate(TRANSFER=factor(TRANSFER,levels = c(0,1,2),labels = c('No','Hospital','Rehab Unit'))) %>%
  mutate(PRIMARYINSURER=factor(PRIMARYINSURER,levels=c(1,2,3,4,5,6),
                               labels = c('Medicare','Medicaid','Commercial', 'Military/VA',
                                          'Non US Insurance','Self Pay'))) %>%
  mutate(LIVINGSTATUS=factor(LIVINGSTATUS,levels=c(1,2,3),
                             labels=c('Home','Nursing home','Homeless'))) %>%
  mutate(PREOP_FUNCSTATUS=factor(PREOP_FUNCSTATUS,levels = c(0,1,2,3,4),
                                 labels = c('Full','Light work','Self care','Assisted care',
                                            'Bed bound'))) %>%
  mutate(PRIOR_CVD = factor(PRIOR_CVD,levels =c(0,1,2,3),labels = c('No','Yes','Yes','Yes'))) %>%
  mutate(PRIOR_CAD = factor(PRIOR_CAD,levels =c(0,1,2,3,4,5),
                            labels = c('No','Yes','Yes','Yes','Yes','Yes'))) %>%
  mutate(PRIOR_CHF = factor(PRIOR_CHF,levels =c(0,1,2,3,4),
                            labels = c('No','Yes','Yes','Yes','Yes'))) %>%
  mutate(COPD = factor(COPD,levels = c(0,1,2,3),labels = c('No','Yes','Yes','Yes'))) %>%
  mutate(DIABETES=factor(DIABETES,levels = c(0,1,2,3),labels = c('No','Yes','Yes','Yes'))) %>%
  mutate(PREOP_DIALYSIS=factor(PREOP_DIALYSIS,levels=c(0,1,2),labels=c('No','Yes','Yes'))) %>%
  mutate(HTN=factor(HTN,levels = c(0,1,2,3),labels = c('No','Yes','Yes','Yes'))) %>%
  mutate(PREOP_SMOKING=factor(PREOP_SMOKING,levels=c(0,1,2),labels=c('No','Yes','Yes'))) %>%
  mutate(PRIOR_CABG = factor(PRIOR_CABG,levels = c(0,1,2),labels = c('No','Yes','Yes'))) %>%
  mutate(PRIOR_PCI = factor(PRIOR_PCI,levels = c(0,1,2),labels = c('No','Yes','Yes'))) %>%
  mutate(PRIOR_ANEURREP = factor(PRIOR_ANEURREP,levels =c(0,1,2,3,4,5),
                                 labels = c('No','Yes','Yes','Yes','Yes','Yes'))) %>%
  mutate(STRESS = factor(STRESS,levels =c(0,1,2,3,4),
```

```
                              labels = c('No','Yes','Yes','Yes','Yes'))) %>%
  mutate(DC_ASA = factor(DC_ASA,levels = c(0,1,2,3),labels = c('No','Yes','No','No'))) %>%
  mutate(DC_P2Y = factor(DC_P2Y,levels =c(0,1,2,3,4,5,6,7),
                         labels = c('No','Yes','Yes','Yes','Yes','Yes','No','No'))) %>%
  mutate(DC_STATIN = factor(DC_STATIN,levels = c(0,1,2,3),labels = c('No','Yes','No','No'))) %>%
  ## merge PREOP_CREAT, R_CR_PRESENT
  mutate(PREOP_CREAT = coalesce(PREOP_CREAT,R_CR_PRESENT))


## ------------- variables cleaning, operative variables----------

TEVAR = TEVAR %>% mutate(
  PRESENTATION = factor(PRESENTATION,levels = c(0,1),labels = c('Asymptomatic','Symptomatic')),
  extent = ifelse((PROXZONE_DISEASE %in% c(2,3)) & DISTZONE_DISEASE<6, 1,
             ifelse((PROXZONE_DISEASE %in% c(2,3)) & DISTZONE_DISEASE>=8, 2,
                ifelse((PROXZONE_DISEASE %in% c(4,5)) & DISTZONE_DISEASE>8, 3,
                   ifelse((PROXZONE_DISEASE <=7)& DISTZONE_DISEASE>=8,4,
                      ifelse((PROXZONE_DISEASE %in% c(4,5)) & DISTZONE_DISEASE<9, 5,
                         ifelse( PROXZONE_DISEASE<=8 & DISTZONE_DISEASE>=9, 6,NA))
  mutate(extent = factor(extent,levels=c(1,2,3,4,5,6),
                         labels = c("Type 1 TAAA", "Type 2 TAAA" ,"Type 3 TAAA" ,"Type 4 TAAA" ,
                                    "Type 5 TAAA" ,"Juxtarenal AAA"))) %>%
  mutate(PROXZONE_DISEASE=factor(PROXZONE_DISEASE)) %>%
  mutate(DISTZONE_DISEASE=
           factor(DISTZONE_DISEASE, levels = c(0:15),
                  labels = c(0,1,2,3,4,5,6,7,8,9,'10R','10L','10B','11R','11L','11B')))




TEVAR = TEVAR %>%
  mutate(PATHOLOGY=factor(PATHOLOGY,levels=c(1,2,3,5,6,7),
                          labels = c('Aneurysm','Dissection','Aneurysm from dissection',
                                     'PAU/IMH','PAU/IMH','PAU/IMH'))) %>%
  mutate(PRIOR_AORSURG=factor(PRIOR_AORSURG,levels=c(0,1,2,3,4),
                          labels=c('None','Open','Endo','Both','Other'))) %>%
  mutate(PATHOLOGY_ANEURYSM_TYPE=
           factor(PATHOLOGY_ANEURYSM_TYPE,levels = c(1,2,3,4,5),
                  labels = c('Degenerative, fusiform','Degenerative, saccular','Anastomotic',
                             'Prior trauma','Intercostal or visceral patch'))) %>%
  mutate(PATHOLOGY_DISSECT_TYPE=
           factor(PATHOLOGY_DISSECT_TYPE,levels = c(1,2),
                  labels = c('Acute, <= 30 days','Chronic, >30 days'))) %>%
  mutate(GENHIST = factor(GENHIST,levels = c(0,1,2,3,4,5),
                          labels = c('None','Marfans','Ehlers-Danlos','Loeys-Dietz',
                                     'Non-specific','Other'))) %>%
  mutate(ANESTHESIA=factor(ANESTHESIA,levels = c(1,2,3),labels = c('Local','Regional','General')))%>%
  mutate(IVUSTEE=factor(IVUSTEE,levels = c(0:5),labels = c('No','IVUS','TEE','Both','No','IVUS')))%>%
  ## merge ACCESS_L and ACCESS_R
  mutate(ACCESS = ifelse(ACCESS_L == 1 & ACCESS_R ==1, 'Percutaneous','Open')) %>%
  mutate(ACCESS=factor(ACCESS)) %>%
  mutate(ARMNECK_ACCESS=factor(ARMNECK_ACCESS,levels =c(0,1,2,3),
                               labels = c('No','Yes','Yes','Yes')))%>%
  mutate(AORDEV_NUM=factor(AORDEV_NUM))%>%
  mutate(AORDEV_CMOD=factor(AORDEV_CMOD,levels=c(0,1),labels=c('No','Yes'))) %>%
```

```r
    mutate(STAGEDAORTRT=factor(STAGEDAORTRT,levels=c(0,1),labels=c('No','Yes'))) %>%
    ## DEV_GTYPE: merge DEV1_GTYPE, DEV2_GTYPE, DEV3_GTYPE
    mutate(DEV_GTYPE =  case_when(DEV1_GTYPE==2|DEV2_GTYPE==2|DEV3_GTYPE==2 ~ 'Custom',
                                  DEV1_GTYPE==3|DEV2_GTYPE==3|DEV3_GTYPE==3 ~ 'Physician modified',
                                  TRUE ~ 'Standard')) %>%
    mutate(DEV1_GTYPE=factor(DEV1_GTYPE,levels=c(1,2,3),
                             labels=c('Standard','Custom','Physician modified'))) %>%
    mutate(DEV2_GTYPE=factor(DEV2_GTYPE,levels=c(1,2,3),
                             labels=c('Standard','Custom','Physician modified'))) %>%
    mutate(DEV3_GTYPE=factor(DEV3_GTYPE,levels=c(1,2,3),
                             labels=c('Standard','Custom','Physician modified')))

TEVAR = TEVAR %>%
  mutate(ILIACDEV_END_R= factor(ILIACDEV_END_R, levels = c(0,1,2,3),
                                labels = c('None','Common',
                                           'External,Intended','External, Unintended'))) %>%
  mutate(ILIACDEV_END_L= factor(ILIACDEV_END_L, levels = c(0,1,2,3),
                                labels = c('None','Common',
                                           'External,Intended','External, Unintended'))) %>%
  mutate(BRANCH_STAGED=factor(BRANCH_STAGED,levels=c(0,1),labels=c('No','Yes'))) %>%
  mutate(BRANCH_LSUB=factor(BRANCH_LSUB,levels=c(0,1),labels=c('No','Yes'))) %>%
  mutate(BRANCH_CELIAC=factor(BRANCH_CELIAC,levels=c(0,1),labels=c('No','Yes'))) %>%
  mutate(BRANCH_SMA=factor(BRANCH_SMA,levels=c(0,1),labels=c('No','Yes'))) %>%
  mutate(BRANCH_RRENAL=factor(BRANCH_RRENAL,levels=c(0,1),labels=c('No','Yes'))) %>%
  mutate(BRANCH_LRENAL=factor(BRANCH_LRENAL,levels=c(0,1),labels=c('No','Yes'))) %>%
  mutate(BRANCH_INNO_POST=factor(BRANCH_INNO_POST,levels=c(1,2,3),
                                 labels=c('Patent','Stenosis/Partial Coverage > 50%','Occluded'))) %>%
  mutate(BRANCH_LSUB_VERTPAT=
           factor(BRANCH_LSUB_VERTPAT,levels=c(1:7),
                  labels=c('Patent bilat','Patent bilat, L dominant','Patent bilat, R dominant',
                           'Occluded L,  patent R','Occluded R, patent L','Occluded bilat',
                           'Not imaged'))) %>%
  mutate(ANESTHESIA_GEN_TIMEEXT=factor(ANESTHESIA_GEN_TIMEEXT,levels=c(1,2,4,5),
                                       labels=c('In OR','<12 hrs','12-24 hrs','>24 hrs'))) %>%
  mutate(POSTOP_SPINALDRAIN=factor(POSTOP_SPINALDRAIN,levels=c(0,1,2,3),
                                   labels=c('No','Yes','Yes','Yes')))

## ------------- BRANCH_TRT filtering----------
TEVAR<-TEVAR %>%
  mutate(lrenal = ifelse(BRANCH_LRENAL_TRT %in% c(0,6,7), 0,
                         ifelse(BRANCH_LRENAL_TRT %in% c(9,10,11,12,13,14), 1,
                                ifelse(BRANCH_LRENAL_TRT %in% c(1,2,3,4), 2,
                                       ifelse(BRANCH_LRENAL_TRT == 8, 3,NA)))),
         rrenal = ifelse(BRANCH_RRENAL_TRT %in% c(0,6,7), 0,
                         ifelse(BRANCH_LRENAL_TRT %in% c(9,10,11,12,13,14), 1,
                                ifelse(BRANCH_LRENAL_TRT %in% c(1,2,3,4), 2,
                                       ifelse(BRANCH_LRENAL_TRT == 8, 3,NA)))),
         sma = ifelse(BRANCH_SMA_TRT %in% c(0,6,7), 0,
                      ifelse(BRANCH_SMA_TRT %in% c(9,10,11,12,13,14), 1,
                             ifelse(BRANCH_SMA_TRT %in% c(1,2,3,4), 2,
                                    ifelse(BRANCH_SMA_TRT == 8,3,NA)))),
         celiac = ifelse(BRANCH_CELIAC_TRT %in% c(0,6,7), 0,
                         ifelse(BRANCH_CELIAC_TRT %in% c(9,10,11,12,13,14), 1,
```

```r
                                   ifelse(BRANCH_CELIAC_TRT %in% c(1,2,3,4), 2,
                                          ifelse(BRANCH_CELIAC_TRT == 8,3,NA)))),
           lsub = ifelse(BRANCH_LSUB_TRT %in% c(0,6,7), 0,
                         ifelse(BRANCH_LSUB_TRT %in% c(9,10,11,12,13,14), 1,
                                ifelse(BRANCH_LSUB_TRT %in% c(1,2,3,4), 2,
                                       ifelse(BRANCH_LSUB_TRT == 8,3,NA)))))


## ------------- FBVAR filtering----------
TEVAR = TEVAR %>%
  filter(lrenal == 1 | rrenal== 1 | sma== 1 | celiac== 1)


## ------------- vessel treatment status ----------

TEVAR = TEVAR %>%
  mutate(TREATED_LRENAL=case_when(lrenal==1|lrenal==3 ~ TRUE, TRUE~FALSE)) %>%
  mutate(TREATED_RRENAL=case_when(rrenal==1|rrenal==3 ~ TRUE, TRUE~FALSE)) %>%
  mutate(TREATED_SMA=case_when(sma==1|sma==3 ~ TRUE, TRUE~FALSE)) %>%
  mutate(TREATED_CELIAC=case_when(celiac==1|celiac==3 ~ TRUE, TRUE~FALSE)) %>%
  mutate(TREATED_LSUB=case_when(lsub==1|lsub==3 ~ TRUE, TRUE~FALSE)) %>%


##### Number of treated branches: 4,3,2,1
  mutate(NUM_TREATED_BRANCHES =
           TREATED_LRENAL+TREATED_RRENAL+TREATED_SMA+TREATED_CELIAC) %>%
  mutate(NUM_TREATED_BRANCHES=factor(NUM_TREATED_BRANCHES)) %>%


##### Number of treated renals: 2,1,0
  mutate(NUM_TREATED_RENALS = TREATED_LRENAL+TREATED_RRENAL) %>%
  mutate(NUM_TREATED_RENALS=factor(NUM_TREATED_RENALS)) %>%


##### RENAL occluded/covered: yes or no
  mutate(OCCLUDED_RENAL = case_when(lrenal == 2|rrenal == 2 ~ TRUE,TRUE ~ FALSE)) %>%
##### SMA occluded/covered: yes or no
  mutate(OCCLUDED_SMA = case_when(sma == 2 ~ TRUE,TRUE ~ FALSE)) %>%
##### CELIAC occluded/covered: yes or no
  mutate(OCCLUDED_CELIAC = case_when(celiac == 2 ~ TRUE,TRUE ~ FALSE))

## ------------- set BRANCHES labels----------
TEVAR = TEVAR %>%
  mutate(lrenal=factor(lrenal,levels=c(0,1,2,3),
                       labels=c('None','Scallop/Fen/Branch/Chimney','Occluded/Covered','Scallop/Fen/Bran
  mutate(rrenal=factor(rrenal,levels=c(0,1,2,3),
                       labels=c('None','Scallop/Fen/Branch/Chimney','Occluded/Covered','Scallop/Fen/Bran
  mutate(sma=factor(sma,levels=c(0,1,2,3),
                    labels=c('None','Scallop/Fen/Branch/Chimney','Occluded/Covered','Scallop/Fen/Bran
  mutate(celiac=factor(celiac,levels=c(0,1,2,3),
                       labels=c('None','Scallop/Fen/Branch/Chimney','Occluded/Covered','Scallop/Fen/Bran
  mutate(lsub=factor(lsub,levels=c(0,1,2,3),
                     labels=c('None','Scallop/Fen/Branch/Chimney','Occluded/Covered','Scallop/Fen/Bra

## ------------- variables cleaning, outcomes----------
TEVAR = TEVAR %>%
  mutate(DEAD=factor(DEAD,levels=c(0,1),labels = c(FALSE,TRUE))) %>%
  mutate(POSTOP_LOS = case_when(POSTOP_LOS>7 ~ '> 7',
```

```
                                    TRUE ~ '<= 7')) %>%
mutate(AORDEV_TECHSUCC=factor(AORDEV_TECHSUCC,levels=c(0,1),labels=c('No','Yes'))) %>%
mutate(CONVTOOPEN=factor(CONVTOOPEN,levels=c(0,1),labels=c('No','Yes'))) %>%
mutate(LEAKATCOMP_NONE=factor(LEAKATCOMP_NONE,levels=c(0,1),labels = c('No','Yes'))) %>%
mutate(POSTOP_VASO=factor(POSTOP_VASO,levels=c(0,1,2,3),labels=c('No','Yes','Yes','Yes'))) %>%
mutate(POSTOP_COMPLICATIONS=factor(POSTOP_COMPLICATIONS,levels=c(0,1),labels=c('No','Yes'))) %>%
# ACCESS_COMPLICATION
mutate (ACCESS_COMPLICATION=
        case_when(ACCESS_HEMATOMA_R!=0 | ACCESS_HEMATOMA_L!=0 |
                      ACCESS_OCCLUSION_R !=0 | ACCESS_OCCLUSION_L !=0 ~ "Yes",
                  ACCESS_HEMATOMA_R==0 & ACCESS_HEMATOMA_L==0 &
                      ACCESS_OCCLUSION_R ==0 & ACCESS_OCCLUSION_L ==0 ~ "No")) %>%
## POSTOP_AH: merge POSTOP_MI,POSTOP_DYSRHYTHMIA,POSTOP_CHF
mutate(POSTOP_MI=factor(POSTOP_MI,levels=c(0,1,2),labels=c('No','Yes','Yes'))) %>%
mutate(POSTOP_DYSRHYTHMIA=factor(POSTOP_DYSRHYTHMIA,levels=c(0,1),labels=c('No','Yes'))) %>%
mutate(POSTOP_CHF=factor(POSTOP_CHF,levels=c(0,1),labels=c('No','Yes'))) %>%
mutate(POSTOP_AH=case_when(
   POSTOP_MI=='Yes'|POSTOP_DYSRHYTHMIA=='Yes'|POSTOP_CHF=='Yes' ~ "Yes",
   POSTOP_MI=='No'|POSTOP_DYSRHYTHMIA=='No'|POSTOP_CHF=='No' ~ "No")) %>%

mutate(POSTOP_CEREBROSX = case_when(POSTOP_CEREBROSX >0 ~ 'Yes',
                                    POSTOP_CEREBROSX ==0 ~ 'No')) %>%
mutate(POSTOP_RESPIRATORY=factor(POSTOP_RESPIRATORY,levels = c(0:3),
                                 labels = c('No','Yes','Yes','Yes'))) %>%
mutate(POSTOP_DIALYSIS=factor(POSTOP_DIALYSIS,levels = c(0:2),
                              labels = c('No','Yes','Yes'))) %>%
mutate(POSTOP_ARMEMBO=factor(POSTOP_ARMEMBO,levels = c(0:5),
                             labels = c('No','Yes','Yes','Yes','Yes','Yes'))) %>%
mutate(POSTOP_LEGEMBO=factor(POSTOP_LEGEMBO,levels = c(0:5),
                             labels = c('No','Yes','Yes','Yes','Yes','Yes'))) %>%
mutate(POSTOP_LEGCOMPART=factor(POSTOP_LEGCOMPART,levels = c(0:4),
                                labels = c('No','Yes','Yes','Yes','Yes'))) %>%
mutate(POSTOP_RENALISCH=factor(POSTOP_RENALISCH,levels = c(0:3),
                               labels = c('No','Yes','Yes','Yes'))) %>%
mutate(POSTOP_SPINAL_ISCHEMIA=factor(POSTOP_SPINAL_ISCHEMIA,levels = c(0:2),
                                     labels = c('No','Yes','Yes'))) %>%
## RETX_R_RTOR: merge RETX, R_RTOR
mutate(RETX=factor(RETX,levels = c(0:2),labels = c('No','Yes','Yes'))) %>%
mutate(R_RTOR=factor(R_RTOR,levels = c(0:5),labels = c('No','Yes','Yes','Yes','Yes','Yes'))) %>%
mutate (RETX_R_RTOR=case_when(RETX=="Yes"|R_RTOR=="Yes" ~ "Yes",
                              RETX=="No" & R_RTOR=="No" ~ "No",
                              RETX=="No"&is.na(R_RTOR) ~ "No",
                              R_RTOR=="No"&is.na(RETX) ~ "No",
                              is.na(R_RTOR) & is.na(RETX) ~ NA_character_)) %>%
mutate(DC_STATUS=factor(DC_STATUS,levels = c(1:6),
                    labels = c('Home','Rehab Unit','Nursing Home',
                               'Dead','Other Hospital','Homeless'))) %>%
## BRANCH_XXX_POST
mutate(BRANCH_LSUB_POST=factor(BRANCH_LSUB_POST,levels=c(1,2,3),
                               labels=c('No','Yes','Yes'))) %>%
mutate(BRANCH_CELIAC_POST=factor(BRANCH_CELIAC_POST,levels=c(1,2,3),
                                 labels=c('No','Yes','Yes'))) %>%
mutate(BRANCH_SMA_POST=factor(BRANCH_SMA_POST,levels=c(1,2,3),
```

```r
                                             labels=c('No','Yes','Yes'))) %>%
   mutate(BRANCH_RRENAL_POST=factor(BRANCH_RRENAL_POST,levels=c(1,2,3),
                                             labels=c('No','Yes','Yes'))) %>%
   mutate(BRANCH_LRENAL_POST=factor(BRANCH_LRENAL_POST,levels=c(1,2,3),
                                             labels=c('No','Yes','Yes'))) %>%
   mutate(BRANCH_RCOMILI_POST=factor(BRANCH_RCOMILI_POST,levels=c(1,2,3),
                                             labels=c('No','Yes','Yes'))) %>%
   mutate(BRANCH_LCOMILI_POST=factor(BRANCH_LCOMILI_POST,levels=c(1,2,3),
                                             labels=c('No','Yes','Yes'))) %>%


   ## BRANCH_POST: merge BRANCH_XXX_POST
   mutate(BRANCH_POST=case_when(BRANCH_LSUB_POST=="Yes"|BRANCH_CELIAC_POST=="Yes"|BRANCH_SMA_POST=="Yes"

# TEVAR = TEVAR %>%
#   mutate(R_PREOP_AMBUL = factor(R_PREOP_AMBUL,levels = c(1,2,3,4),
#                                 labels=c("Amb","Amb w/ Assistance","Wheelchair","Bedridden"))) %>%
#   mutate(SURGYEAR=factor(SURGYEAR)) %>%
#   mutate(PRIOR_AORSURG_OPENLOC1=factor(PRIOR_AORSURG_OPENLOC1,
#                                 levels=c(0,1),labels=c('No','Yes'))) %>%
#   mutate(PRIOR_AORSURG_OPENLOC2=factor(PRIOR_AORSURG_OPENLOC2,
#                                 levels=c(0,1),labels=c('No','Yes'))) %>%
#   mutate(PRIOR_AORSURG_OPENLOC3=factor(PRIOR_AORSURG_OPENLOC3,
#                                 levels=c(0,1),labels=c('No','Yes'))) %>%
#   mutate(PRIOR_AORSURG_OPENLOC4=factor(PRIOR_AORSURG_OPENLOC4,
#                                 levels=c(0,1),labels=c('No','Yes'))) %>%
#   mutate(PRIOR_AORSURG_ENDOLOC1=factor(PRIOR_AORSURG_ENDOLOC1,
#                                 levels=c(0,1),labels=c('No','Yes'))) %>%
#   mutate(PRIOR_AORSURG_ENDOLOC2=factor(PRIOR_AORSURG_ENDOLOC2,
#                                 levels=c(0,1),labels=c('No','Yes'))) %>%
#   mutate(PRIOR_AORSURG_ENDOLOC3=factor(PRIOR_AORSURG_ENDOLOC3,
#                                 levels=c(0,1),labels=c('No','Yes'))) %>%
#   mutate(PRIOR_AORSURG_ENDOLOC4=factor(PRIOR_AORSURG_ENDOLOC4,
#                                 levels=c(0,1),labels=c('No','Yes'))) %>%
#   mutate(ARMNECK_ACCESS_LOC=factor(ARMNECK_ACCESS_LOC,levels =c(1:7),
#                                 labels = c('Right arm','Left arm','Right axillary',
#                                            'Left axillary','Right carotid',
#                                            'Left carotid','Multiple'))) %>%
#   ## retired branch variables
#   mutate(R_CELIAC=factor(R_CELIAC,levels=c(0:9),
#                       labels=c('Patent, no intervention','Chronically Occluded',
#                                'Purposely Occluded','De-branch','Stent Only','Chimney',
#                                'Fen/scallop Only','Stented-fen',
#                                'Fenestrated Stentgraft Branch (Branched TEVAR)',
#                                'Side-arm Stent-graft Branch'))) %>%
#   mutate(R_LT_RENAL=factor(R_LT_RENAL,levels=c(0:9),
#                       labels=c('Patent, no intervention','Chronically Occluded',
#                                'Purposely Occluded','De-branch','Stent Only','Chimney',
#                                'Fen/scallop Only','Stented-fen',
#                                'Fenestrated Stentgraft Branch (Branched TEVAR)',
#                                'Side-arm Stent-graft Branch'))) %>%
#   mutate(R_RT_RENAL=factor(R_RT_RENAL,levels=c(0:9),
#                       labels=c('Patent, no intervention','Chronically Occluded',
#                                'Purposely Occluded','De-branch','Stent Only','Chimney',
```

```
#                                            'Fen/scallop Only','Stented-fen',
#                                            'Fenestrated Stentgraft Branch (Branched TEVAR)',
#                                            'Side-arm Stent-graft Branch'))) %>%
#   mutate(R_SMA=factor(R_SMA,levels=c(0:9),
#                           labels=c('Patent, no intervention','Chronically Occluded',
#                                            'Purposely Occluded','De-branch','Stent Only','Chimney',
#                                            'Fen/scallop Only','Stented-fen',
#                                            'Fenestrated Stentgraft Branch (Branched TEVAR)',
#                                            'Side-arm Stent-graft Branch')))
#
# ## ------------- DEV_PROXZONE and DEV_DISTZONE filtering----------
# TEVAR<-TEVAR %>%
#   mutate(distal_seal= DEV1_DISTZONE,prox_seal= DEV1_PROXZONE)
#
# for (i in 1:nrow(TEVAR)){
#   if ((!(is.na(TEVAR$DEV2_DISTZONE[i])) && (!is.na(TEVAR$DEV2_DISTZONE[i]>TEVAR$DEV1_DISTZONE[i]))&&
#        (TEVAR$DEV2_DISTZONE[i]>TEVAR$DEV1_DISTZONE[i]))) {
#     TEVAR$distal_seal[i] = TEVAR$DEV2_DISTZONE[i]}
#   else if((is.na(TEVAR$DEV1_DISTZONE[i])) && (!(is.na(TEVAR$DEV2_DISTZONE[i])))){
#        TEVAR$distal_seal[i] = TEVAR$DEV2_DISTZONE[i]}
#   if ((!(is.na(TEVAR$DEV3_DISTZONE[i])) && (!is.na(TEVAR$DEV3_DISTZONE[i]>TEVAR$DEV2_DISTZONE[i]))&&
#        (TEVAR$DEV3_DISTZONE[i]>TEVAR$DEV2_DISTZONE[i]))) {
#     TEVAR$distal_seal[i] = TEVAR$DEV3_DISTZONE[i]}
#   if ((!(is.na(TEVAR$DEV4_DISTZONE[i])) && (!is.na(TEVAR$DEV4_DISTZONE[i]>TEVAR$DEV3_DISTZONE[i]))&&
#        (TEVAR$DEV4_DISTZONE[i]>TEVAR$DEV3_DISTZONE[i]))) {
#     TEVAR$distal_seal[i] = TEVAR$DEV4_DISTZONE[i]}
#   if ((!(is.na(TEVAR$DEV5_DISTZONE[i])) && (!is.na(TEVAR$DEV5_DISTZONE[i]>TEVAR$DEV4_DISTZONE[i]))&&
#        (TEVAR$DEV5_DISTZONE[i]>TEVAR$DEV4_DISTZONE[i]))) {
#     TEVAR$distal_seal[i] = TEVAR$DEV5_DISTZONE[i]}
#   if ((!(is.na(TEVAR$DEV6_DISTZONE[i])) && (!is.na(TEVAR$DEV6_DISTZONE[i]>TEVAR$DEV5_DISTZONE[i]))&&
#        (TEVAR$DEV6_DISTZONE[i]>TEVAR$DEV5_DISTZONE[i]))) {
#     TEVAR$distal_seal[i] = TEVAR$DEV6_DISTZONE[i]}
#   }
#
# for (i in 1:nrow(TEVAR)){
#   if ((!(is.na(TEVAR$DEV2_PROXZONE[i])) && (!is.na(TEVAR$DEV2_PROXZONE[i]<TEVAR$DEV1_PROXZONE[i]))&&
#        (TEVAR$DEV2_PROXZONE[i]<TEVAR$DEV1_PROXZONE[i]))) {
#     TEVAR$prox_seal[i] = TEVAR$DEV2_PROXZONE[i]}
#   else if((is.na(TEVAR$DEV1_PROXZONE[i])) && (!(is.na(TEVAR$DEV2_PROXZONE[i])))){
#        TEVAR$prox_seal[i] = TEVAR$DEV2_PROXZONE[i]}
#   if ((!(is.na(TEVAR$DEV3_PROXZONE[i])) && (!is.na(TEVAR$DEV3_PROXZONE[i]<TEVAR$DEV2_PROXZONE[i]))&&
#        (TEVAR$DEV3_PROXZONE[i]<TEVAR$DEV2_PROXZONE[i]))) {
#     TEVAR$prox_seal[i] = TEVAR$DEV3_PROXZONE[i]}
#   if ((!(is.na(TEVAR$DEV4_PROXZONE[i])) && (!is.na(TEVAR$DEV4_PROXZONE[i]<TEVAR$DEV3_PROXZONE[i]))&&
#        (TEVAR$DEV4_PROXZONE[i]<TEVAR$DEV3_PROXZONE[i]))) {
#     TEVAR$prox_seal[i] = TEVAR$DEV4_PROXZONE[i]}
#   if ((!(is.na(TEVAR$DEV5_PROXZONE[i])) && (!is.na(TEVAR$DEV5_PROXZONE[i]<TEVAR$DEV4_PROXZONE[i]))&&
#        (TEVAR$DEV5_PROXZONE[i]<TEVAR$DEV4_PROXZONE[i]))) {
#     TEVAR$prox_seal[i] = TEVAR$DEV5_PROXZONE[i]}
#   if ((!(is.na(TEVAR$DEV6_PROXZONE[i])) && (!is.na(TEVAR$DEV6_PROXZONE[i]<TEVAR$DEV5_PROXZONE[i]))&&
#        (TEVAR$DEV6_PROXZONE[i]<TEVAR$DEV5_PROXZONE[i]))) {
#     TEVAR$prox_seal[i] = TEVAR$DEV6_PROXZONE[i]}
# }
```

```r
#

## ------------ variables cleaning, volume variables----------
TEVAR = TEVAR %>%
  mutate(REGIONID=factor(REGIONID)) %>%
  mutate(CENTERID=factor(CENTERID)) %>%
  mutate(PHYSICIANID=factor(PHYSICIANID))


## ------------variables selection----------
FBVAR = TEVAR %>% select(
  PRESENTATION,
  # Patient demographic and co-morbidities
  AGE, AGECAT, GENDER, ETHNICITY, RACE, TRANSFER, PRIMARYINSURER, LIVINGSTATUS, PREOP_FUNCSTATUS, PRIOR
  # Operative Variables
  PRIOR_AORSURG, PATHOLOGY, PREOP_MAXAAADIA, URGENCY, PATHOLOGY_ANEURYSM_TYPE, PATHOLOGY_DISSECT_TYPE,
  # Outcomes
  DEAD, PROC_SURVIVALDAYS, LTF_NUM_REINT,TOTAL_LOS, POSTOP_LOS, AORDEV_TECHSUCC, CONVTOOPEN, LEAKATCOMP
  # Volume variables,
  REGIONID,CENTERID,PHYSICIANID)

## ------------ store as new dataset 'FBVAR' to ensure easier access for modelling----------
write.csv(FBVAR,path_lily[5])
#write.csv(FBVAR,path_jenn[5])
```