

# 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, IDE\_OTHER. Merge by these variables.

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
- extent: type of TAAA based on certain criteria calculated by PROXZONE\_DISEASE and DISTZONE\_DISEASE. *A lot of missing?*
- ILIACDEV\_END\_R, ILIACDEV\_END\_L: *a lot of missing, merge LR?*
- 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`, merge with retired `R_LT_RENAL`
- **rrenal:** re-leveled `BRANCH_RRENAL_TRT`, merge with retired `R_RT_RENAL`
- **sma:** re-leveled `BRANCH_SMA_TRT`, merge with retired `R_SMA`
- **celiac:** re-leveled `BRANCH_CELIAC_TRT`, merge with retired `R_CELIAC`
- **lsub:** re-leveled `BRANCH_LSUB_TRT`, merge with retired `R_L_SUBCLAV`

Retired levels: 0 = Patent, no intervention, 1 = Chronically Occluded, 2 = Purposely Occluded, 3 = De-branch, 4 = Stent Only, 5 = Chimney, 6 = Fen/scallop Only, 7 = Stented-fen, 8 = Fenestrated Stentgraft Branch (Branched TEVAR), 9 = Side-arm Stent-graft Branch

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

*How to map new and old levels?*

### Retired variables

- **GENE:** merge `GENHIST` with retired `R_GENHIST`
- **POSTOP\_SPINALDRAIN:** merge `POSTOP_SPINALDRAIN` with retired `R_SPINAL_DRAIN`
- **R\_DISTATTZONE:** Distal Attachment Zone *Don't know which is the current variable*
- **R\_GDPROXIMAL:** Graft Diameter Proximal *Don't know which is the current variable*
- **R\_GRFTCONFIG:** Graft Configuration *Don't know which is the current variable*
- **R\_PRATTZONE:** Prox. Attachment Zone *Don't know which is the current variable*

Among the whole dataset, we filtered out 3510 FBVAR patients for our study.

### Outcomes

- **R\_ENDOLEAK\_AT\_COMPLETION, R\_POSTOP\_HEMATOMA, R\_POSTOP\_SITEOCC:** changed to Yes/No scale
- **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`

### Retired variables

- **R\_ENDOLEAK\_AT\_COMPLETION:** *many LEAKATCOM\_XXX variables, don't know how to pick*
- **R\_POSTOP\_HEMATOMA:** corresponding to retired `ACCESS_HEMATOMA_R` `ACCESS_HEMATOMA_L` Merge LR?
- **R\_POSTOP\_SITEOCC:** corresponding to retired `ACCESS_OCCLUSION_R` + `ACCESS_OCCLUSION_L` Merge LR?
- **R\_POSTOP\_SSI:** corresponding to retired `ACCESS_INFECTION_R` + `ACCESS_INFECTION_L` Merge LR?

**a lot of variables related to Ischemia?**

- R\_POSTOP\_BOWELISCH: Bowel Ischemia
- R\_LE\_ISCH: LE Ischemia
- R\_POSTOP\_RENAL: change of renal function

**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 separate csv file, so that we could use the cleaned dataset in the future modeling.

## Code Appendix

```
knitr::opts_chunk$set(echo = FALSE,message = FALSE,warning = FALSE)
library(tidyverse)
library(table1)
library(Hmisc)
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/jennifercci/Desktop/Comparative-analysis-of-treatments-of-CAA'
path_jenn = c(
  # "/Users/jennifercci/Desktop/Comparative-analysis-of-treatments-of-CAA/TEVAR_International_20210712/TE
  # "/Users/jennifercci/Desktop/Comparative-analysis-of-treatments-of-CAA/TEVAR_International_20210712/TE
  # "/Users/jennifercci/Desktop/Comparative-analysis-of-treatments-of-CAA/TEVAR_International_20210901/TE
  # "/Users/jennifercci/Desktop/Comparative-analysis-of-treatments-of-CAA/TEVAR_International_20210901/TE

## ----- 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)

# 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)]

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

```

## ----- 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('<40','40-49','50-59','60-69','70-79','80-89','>89')) %>%
  mutate(PROXZONE_DISEASE=factor(PROXZONE_DISEASE)) %>%
  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(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
  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 %in% c(6,7)) & DISTZONE_DISEASE>=8, 4,
          ifelse((PROXZONE_DISEASE %in% c(4,5)) & DISTZONE_DISEASE<9, 5,
            ifelse( DISTZONE_DISEASE==8 & DISTZONE_DISEASE>8, 6,
              ifelse( DISTZONE_DISEASE==9 & DISTZONE_DISEASE>9,
                7, NA))))))))) %>%
mutate(extent = factor(extent, levels=c(1,2,3,4,5,6,7),
  labels = c("Type 1 TAAA", "Type 2 TAAA", "Type 3 TAAA", "Type 4 TAAA",
    "Type 5 TAAA", "Suprarenal AAA", "Infrarenal/Juxtarenal AAA"))) %>%
## PATHOLOGY
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'))) %>%
## merge GENHIST, R_GENHIST
mutate(GENHIST = factor(GENHIST, levels = c(0,1,2,3,4,5),
  labels = c('None', 'Marfans', 'Ehlers-Danlos', 'Loeys-Dietz',
    'Non-specific', 'Other'))) %>%
mutate(R_GENHIST = factor(R_GENHIST, levels = c(0,1,2,3),
  labels = c('None', 'Marfans', 'Ehlers-Danlos', 'Other'))) %>%
mutate(GENE=coalesce(GENHIST, R_GENHIST)) %>%
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'))) %>%
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', 'For branch treatment',
    'For femoral-brachial wire', 'For both')))%>%
mutate(AORDEV_NUM=factor(AORDEV_NUM))%>%

```

```

mutate(AORDEV_CMED=factor(AORDEV_CMED,levels=c(0,1),labels=c('No','Yes'))) %>%
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(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')) %>%
## POSTOP_SPINALDRAIN
mutate(R_SPINAL_DRAIN=factor(R_SPINAL_DRAIN,levels=c(0,1,2),
                             labels=c('None','Pre-op','Post-op')) %>%
mutate(POSTOP_SPINALDRAIN=factor(POSTOP_SPINALDRAIN,levels=c(0,1,2,3),
                                labels=c('None','Pre-op','Post-op, prophylactic',
                                           'Post-op for spinal ischemia')) %>%
mutate(POSTOP_SPINALDRAIN = coalesce(POSTOP_SPINALDRAIN,R_SPINAL_DRAIN)) %>%
mutate(R_PRATTZONE=factor(R_PRATTZONE)) %>%
mutate(R_GRFTCONFIG=factor(R_GRFTCONFIG,levels=c(1,2,3),
                                labels=c('Standard','Plus Bare Stent','Branched/fenestrated')) %>%
## 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'))

## ----- 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,
        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 | lsub== 1)

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

## ----- variables cleaning, outcomes-----

```



```

TEVAR = TEVAR %>%
  mutate(DEAD=factor(DEAD,levels=c(0,1),labels = c('No','Yes'))) %>%
  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(R_ENDOLEAK_AT_COMPLETION=
    factor(R_ENDOLEAK_AT_COMPLETION,levels=c(0:4),
      labels=c('No','Yes','Yes','Yes','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'))) %>%
  mutate(R_POSTOP_HEMATOMA=factor(R_POSTOP_HEMATOMA,levels=c(0,1,2,3,4),
    labels=c('No','Yes','Yes','Yes','Yes'))) %>%
  mutate(R_POSTOP_SITEOCC=factor(R_POSTOP_SITEOCC,levels=c(0:5),
    labels=c('No','Yes','Yes','Yes','Yes','Yes'))) %>%
  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(R_POSTOP_BOWELISCH=factor(R_POSTOP_BOWELISCH,levels = c(0:2),labels = c('No','Yes','Yes'))) %>%
  mutate(POSTOP_RENALISCH=factor(POSTOP_RENALISCH,levels = c(0:3),
    labels = c('No','Yes','Yes','Yes'))) %>%
  mutate(R_POSTOP_SSI=factor(R_POSTOP_SSI,levels = c(0:4),labels = c('No','Yes','Yes','Yes','Yes'))) %>%
  mutate(POSTOP_SPINAL_ISCHEMIA=factor(POSTOP_SPINAL_ISCHEMIA,levels = c(0:2),
    labels = c('No','Yes','Yes'))) %>%
  mutate(DC_STATUS=factor(DC_STATUS,levels = c(1:6),
    labels = c('Home','Rehab Unit','Nursing Home',
      'Dead','Other Hospital','Homeless'))) %>%
  mutate(R_LE_ISCH=factor(R_LE_ISCH,levels = c(0,1),labels = c('No','Yes'))) %>%
  mutate(R_POSTOP_RENAL=factor(R_POSTOP_RENAL,levels = c(0:5),
    labels = c('No','Yes','Yes','Yes','Yes','Yes'))) %>%

  ## 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")) %>%

  ## 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),
    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"

## 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_))

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(SURGEYEAR=factor(SURGEYEAR)) %>%
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'))))

## ----- 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]}
}

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

## -----variables selection-----
FBVAR = TEVAR %>% select(
  # Patient demographic and co-morbidities
  AGE,AGECAT,GENDER,PROXZONE_DISEASE,URGENCY,ETHNICITY,RACE,TRANSFER,PRIMARYINSURER,
  LIVINGSTATUS,PREOP_FUNCSTATUS,PRIOR_CVD,PRIOR_CAD,PRIOR_CHF,COPD,DIABETES,
  PREOP_DIALYSIS,HTN,PREOP_SMOKING,PRIOR_CABG,PRIOR_PCI,PRIOR_ANEURREP,STRESS,
  PREOP_CREAT,DC_ASA,DC_P2Y,DC_STATIN,PREOP_CREAT,
  lrenal,rrenal,sma,celiac,lsub,
  #operative variables
  PRESENTATION,PRIOR_AORSURG,PRIOR_AORSURG_OPENLOC1,PRIOR_AORSURG_OPENLOC2,PRIOR_AORSURG_OPENLOC3,
  PRIOR_AORSURG_OPENLOC4,PRIOR_AORSURG_ENDOLOC1,PRIOR_AORSURG_ENDOLOC2,

```

```

PRIOR_AORSURG_ENDOLOC3,PRIOR_AORSURG_ENDOLOC4,PATHOLOGY,PREOP_MAXAAADIA,URGENCY,
PATHOLOGY_ANEURYSM_TYPE,PATHOLOGY_DISSECT_TYPE,PATHOLOGY_DISSECT_ONSET_DAYS,GENE,
PROXZONE_DISEASE,DISTZONE_DISEASE,ANESTHESIA,CONTRAST,EBL,FLUOROTIME,INTRAOP_PRBC,
TOTALPROCTIME,IVUSTEE,ACCESS, #ACCESS_R,ACCESS_L,LRGST_SHEATH_SIZE_R,LRGST_SHEATH_SIZE_L,
ARMNECK_ACCESS,ARMNECK_ACCESS_LOC,AORDEV_NUM,
AORDEV_CMED, #AORDEV_TECHSUCC,CONVTOOPEN,R_ENDOLEAK_AT_COMPLETION,STAGEDAORTRT,
DEV_GTYPE, #DEV1_GTYPE,DEV2_GTYPE,DEV3_GTYPE,
extent,
DEV1_GRAFTDET,ILIACDEV_END_R,ILIACDEV_END_L,BRANCH_STAGED,BRANCH_LSUB,
BRANCH_CELIAC,BRANCH_SMA,BRANCH_RRENAL,BRANCH_LRENAL,
#BRANCH_INNO_POST,
#BRANCH_LSUB_POST,BRANCH_CELIAC_POST,BRANCH_SMA_POST,
#BRANCH_RRENAL_POST,BRANCH_LRENAL_POST,BRANCH_RCOMILI_POST,BRANCH_LCOMILI_POST,
#BRANCH_LSUB_VERTPAT,
ANESTHESIA_GEN_TIMEEXT,ICUSTAY,POSTOP_PRBC,POSTOP_SPINALDRAIN,
#POSTOP_VASO,
POSTOP_HIGHCREAT,POSTOP_COMPLICATIONS,#R_POSTOP_HEMATOMA,R_POSTOP_SITEOCC,
#POSTOP_MI,POSTOP_DYSRHYTHMIA,POSTOP_CHF,
POSTOP_CEREBROX,POSTOP_RESPIRATORY,
POSTOP_DIALYSIS,POSTOP_ARMEMBO,POSTOP_LEGEMBO,POSTOP_LEGCOMPART,POSTOP_INTISCH,
R_POSTOP_BOWELISCH,POSTOP_RENALISCH,R_POSTOP_SSI,POSTOP_SPINAL_ISCHEMIA,#RETX,
DC_STATUS,
DC_ASA,DC_P2Y,DC_STATIN,R_CELIAC,#R_CR_PRESENT,
R_DISTATTZONE,R_GDPROXIMAL,R_GRFTCONFIG,
R_LE_ISCH,R_LT_RENAL,R_PRATTZONE,R_POSTOP_RENAL,R_RT_RENAL,#R_RTOR,
R_SMA,#R_SPINAL_DRAIN,
# outcomes
DEAD,PROC_SURVIVALDAYS,TOTAL_LOS,POSTOP_LOS,AORDEV_TECHSUCC,CONVTOOPEN,
R_ENDOLEAK_AT_COMPLETION,BRANCH_POST,#BRANCH_LSUB_POST,BRANCH_CELIAC_POST,BRANCH_SMA_POST,
#BRANCH_RRENAL_POST,BRANCH_LRENAL_POST,BRANCH_RCOMILI_POST,BRANCH_LCOMILI_POST,
ICUSTAY,POSTOP_PRBC,POSTOP_VASO,POSTOP_HIGHCREAT,POSTOP_COMPLICATIONS,R_POSTOP_HEMATOMA,
R_POSTOP_SITEOCC,#POSTOP_MI,POSTOP_DYSRHYTHMIA,POSTOP_CHF,
POSTOP_AH,POSTOP_CEREBROX,POSTOP_RESPIRATORY,
POSTOP_DIALYSIS,POSTOP_ARMEMBO,POSTOP_LEGEMBO,POSTOP_LEGCOMPART,POSTOP_INTISCH,
R_POSTOP_BOWELISCH,POSTOP_RENALISCH,R_POSTOP_SSI,POSTOP_SPINAL_ISCHEMIA,#RETX,
DC_STATUS,R_LE_ISCH,#R_POSTOP_RENAL,#R_RTOR,
RETX_R_RTOR,

# clustering
REGIONID,CENTERID,PHYSICIANID)

## ----- variables labels-----
#var.labels = c(AGE="Age (years)", AGECAT="Age (category)")
#label(FBVAR) = as.list(var.labels[match(names(FBVAR), names(var.labels))])

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

```