Data Preprocessing for the VQI TEVAR Dataset

Jennifer Ci, Thu Vu, Lily Hanyi Wang

Datasets Merging

Compare the data from July 2021, September 2021 and January 2023. Keep the most updated ones by merging LTF and PROC datasets separately.

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

We will work on the PROC and LTF datasets separately. The only variable we want to study in the LTF dataset is the re-intervention variable, `LTF_NUM_REINT`.

For the time-to-event analysis of re-intervention, we work on the LTF dataset and treat it as a multi-event recurrent survival analysis.

For the logistic regression model of re-intervention, we merge it with the PROC dataset to account for the variables as we did for other outcomes. Here we treated re-intervention as a binary outcome.

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.

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

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

Duplicate variables

We checked the duplicate `PATIENTID` in the `PROC` dataset.These patients are transfered a few days or a few months followed the first procedure. We decided to exclude these abnormal data points.

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.