In general, we’ll probably want a folder for analysis and probably a folder for the analytic dataset. Each of those will include code files and the analysis folders can include our output/tables.

Nick,

This is the function I like for creating tables:

<https://www.danieldsjoberg.com/gtsummary/articles/tbl_summary.html>. you can also use tableone <https://cran.r-project.org/web/packages/tableone/vignettes/introduction.html>, but the first one has more flexibility. I have a few examples that I’ve used in my analyses before if that would be helpful.

For the time to event composite outcomes, I realize I have more SAS code on this than R, but hoping you can see the logic and then apply to R.  And then there is a snippet of R code for making a 2y outcome from a version with no time limit.

\*SAS: define treatment failure or death outcome; (death is the binary variable and ttdeath is the time to (“tt”) death in days. Trtfail is treatment failure and tttrtfail is time to trtfailure in days))

  if (death=0 & trtfail=**0**) then deathtrtfail=**0**;

  if (death=1 | trtfail=**1**) then deathtrtfail=**1**;

  if trtfail=**1** then ttdeathtrtfail=tttrtfail; \*trt fail. use trtfail date;

  if trtfail=**0** & death=1 then ttdeathtrtfail=ttdeath; \*no trt fail, yes death. use death date;

  if trtfail=**0** & death=0 then ttdeathtrtfail=min(ttdeath,tttrtfail); \*no trt fail, no death. use earliest censor date;

#R: 2y death (death is the binary variable and ttdeath is the time to (“tt”) death from treatment in days)

data$death\_2y[data $death==0]<-0

data$death\_2y[data$death==1 & data$ttdeath<=365.25\*2]<-1

data$death\_2y[data$death==1 & data$ttdeath>365.25\*2]<-0

data$ttdeath\_2y=pmin(data$ttdeath, 365.25\*2)

As you create new variables, I definitely recommend adding to the data dictionary that Tim started (I prefer it as  stand alone document).