Competing risk regression models in cohort studies with the R package CohortMethod

Kelly Li, Jianxiao Yang, Marc A. Suchard
Department of Biostatistics, University of California, Los Angeles

Background

In survival analysis, competing risks are outcomes that preclude the main study outcome of interest, or alter the probability of experiencing the main outcome(s) [1]. The Fine and Gray model is an extension of the Cox proportional hazards model with subdistribution hazard function (Eq. 1), defined through the instantaneous rate $\lambda_k(t)$ of experiencing the event of interest $D = k$ for subjects given they have survival already to time t without any competing events $K \neq k$ having occurred for observed failure time T [2]:

$$\lambda_k(t) = \lim_{\Delta t \to 0} \frac{\text{Prob}(t \leq T < t + \Delta t, D = k | T \geq t \cup (T < t \cap K \neq k))}{\Delta t}. \quad (1)$$

`CohortMethod` is an R package that performs comparative cohort studies in an observational database in the Observational Medical Outcomes Partnership (OMOP) common data model. `CohortMethod` has the capability of using a large set of covariates to fit propensity models and logistic, Poisson, and Cox regression models. In order to conduct time-to-event analyses, it is necessary to be able to conduct regression on data containing competing risks. The goal of this study is to develop functionality to include the Fine and Gray model as an outcome model to `CohortMethod`, while supporting the ability to trim and match on propensity scores.

Methods

Kawaguchi, Shen, Li, and Suchard (2019) have developed fast and scalable methods for the analysis of large-scale competing risks data. These methods involve using a forward-backward scan algorithm to linearize the computations for log-pseudo likelihood, gradients, and Hessian diagonals, for later use in cyclic coordinate descent. This reduces computational complexity from $\mathcal{O}(N^2)$ to $\mathcal{O}(N)$ in fitting Fine Gray models with N subjects [3].

Previously, `CohortMethod` supported only (conditional) Cox regression. We have developed new function `combineCompetingStudyPopulations` and option `riskId` in `createFitOutcomeModelArgs` to implement Fine Gray model fitting in CohortMethod for both single and multiple analysis frameworks. `combineCompetingStudyPopulations` combines two study populations, one with the outcome of interest, and one with the competing event, to generate a population with information on subjects experiencing either outcome. In the event that both outcomes occur at the same time, there are options to remove these subjects, or prioritize one outcome as the event experienced by the subject. This function is implemented in a multiple analysis framework by specifying the concept ID of the competing risk outcome as the `riskId` in `createFitOutcomeModelArgs`. The regression model is then fitted using OHDSI's `Cyclops` package, which implements the novel forward-backward scan algorithm to estimate the parameters for use in cyclic coordinate descent [4].

Results

We apply `CohortMethod` to study the relative risk of hospitalization with heart failure for new users under angiotensin-converting enzyme (ACE) inhibitors and thiazide diuretics (THZs) treatment in subjects without a prior hospitalization with heart failure or who may experience the competing event of acute myocardial infarction. From the Optum EHR data source, we identify 1,014,618 patients. After applying a large-scale propensity score model, we successfully estimate the hazard ratio under the Fine Gray model among these one-to-one matched new-users of ACE inhibitors and THZs.

To accomplish this using our new extensions, we first use `CohortMethod` to extract the necessary data for our analysis. We specify both our outcome of interest, and the competing risk outcome as `outcomeIds` in `getDbCohortMethodData`. We exclude all first-line hypertension drugs to prevent high correlation between covariates and our treatments.

```
connectionDetails <- createConnectionDetails(dbms = "postgresql",
                                             server = "localhost/ohdsi",
                                             user = "joe",
                                             password = "supersecret")

excludedCovId <- c(904542, 907013, 932745, 942350, 956874, 970250,
                   974166, 978555, 991382, 1305447, 1307046, 1307863,
                   1308216, 1308842, 1309068, 1309799, 1310756, 1313200,
                   1314002, 1314577, 1317640, 1317967, 1318137, 1318853,
                   1319880, 1319998, 1322081, 1326012, 1327978, 1328165,
                   1331235, 1332418, 1334456, 1335471, 1338005, 1340128,
                   1341238, 1341927, 1342439, 1344965, 1345858, 1346686,
                   1346823, 1347384, 1350489, 1351557, 1353766, 1353776,
                   1363053, 1363749, 1367500, 1373225, 1373928, 1386957,
                   1395058, 1398937, 40226742, 40235485)

covSettings <- createDefaultCovariateSettings(
  excludedCovariateConceptIds = excludedCovId,
  addDescendantsToExclude = TRUE)

cmData <- getDbCohortMethodData(
  connectionDetails = connectionDetails,
  cdmDatabaseSchema = cdmDatabaseSchema,
  targetId = 7036,
  comparatorId = 7033,
  outcomeIds = c(7368, 7364),          ### Outcome of interest and competing risk
  exposureDatabaseSchema = "scratch",
  exposureTable = "my_cohorts",
  outcomeDatabaseSchema = "scratch",
  outcomeTable = "my_cohorts",
  cdmVersion = "5",
  firstExposureOnly = FALSE,
  removeDuplicateSubjects = FALSE,
  restrictToCommonPeriod = FALSE,
  washoutPeriod = 0,
  covariateSettings = covSettings
)
```

Next, we define study populations for both the outcome of interest and the competing risk, and we combine them using the new `combineCompetingStudyPopulations` function. We specify for subjects in the data set who experience the outcome of interest and competing risk outcome simultaneously to be removed from the study population. We then fit a propensity model and use the propensity score to performmatching.

```
studyPopOutcome <- createStudyPopulation(cohortMethodData = cmData,
                                         outcomeId = 7368,
                                         firstExposureOnly = FALSE,
                                         riskWindowStart = 0,
                                         riskWindowEnd = 9999)

studyPopRisk <- createStudyPopulation(cohortMethodData = cmData,
                                      outcomeId = 7364,
                                      firstExposureOnly = FALSE,
                                      riskWindowStart = 0,
                                      riskWindowEnd = 9999)

studyPopCombined <- combineCompetingStudyPopulations( ### New function
  mainPopulation = studyPopOutcome,
  competingRiskPopulation = studyPopRisk,
  removeSubjectsWithSimultaneousEvents = TRUE
)

ps <- createPs(cohortMethodData = cmData,
               population = studyPopCombined)

matchedPop <- matchOnPs(population = ps,
                        caliper = 0.2)
```

We fit our Fine Gray outcome model on our one-to-one matched study population:

```
outcome <- fitOutcomeModel(population = matchedPop, ### Updated for multi-type events
                           modelType = "fgr")
outcome
```

```
## Model type: fgr
## Stratified: FALSE
## Use covariates: FALSE
## Use inverse probability of treatment weighting: FALSE
## Status: OK
##
##            Estimate lower .95 upper .95     logRr seLogRr
## treatment  0.942060  0.873580  1.015845 -0.059686  0.0385
```

Conclusions

The `CohortMethod` package now provides the ability to perform comparative cohort studies in an observational database that use the Fine Gray regression model for competing risks. Using the forward-backward scan algorithm, `CohortMethod` calls package `Cyclops` to estimate Fine-Gray parameters in order $\mathcal{O}(N)$ time, rather than $\mathcal{O}(N^2)$, as in other packages of the same purpose, such as `cmprsk`. Thus, its efficiency is along the lines of previously supported survival regression models, such as the Cox proportional hazards models. With this approach, we are able to provide a reliable estimate of Fine Gray regression coefficients in new-user cohort studies with large numbers of observations.

References

[1] Austin PC, Fine JP. Practical recommendations for reporting Fine-Gray model analyses for competing risk data. Statistics in Medicine. 2017;36(27):4391–400.

[2] Fine JP, Gray RJ. A Proportional Hazards Model for the Subdistribution of a Competing Risk. Journal of the American Statistical Association. 1999;94(446):496–509.

[3] Kawaguchi ES, Shen JI, Li G, Suchard MA. A Fast and Scalable Implementation Method for Competing Risks Data with the R Package fastcmprsk. The R Journal. 2020;12(2):163.

[4] Suchard MA, Simpson SE, Zorych I, Ryan P, Madigan D. Massive Parallelization of Serial Inference Algorithms for a Complex Generalized Linear Model. ACM Transactions on Modeling and Computer Simulation. 2013;23(1):1–17.