EP16: Missing Values in Clinical Research: Multiple Imputation

13. Imputation of Survival Data

Nicole Erler

Department of Biostatistics, Erasmus Medical Center

n.erler@erasmusmc.nl

Results from the Literature

In a previous section we saw that the correct conditional distribution for an incomplete covariate x in a proportional hazards model is rather complex:

$$\log p(x \mid T, D, z) = \log p(x \mid z) + D(\beta_x x + \beta_z z) - H_0(T) \exp(\beta_x x + \beta_z z) + const.$$

White and Royston (2009) investigated how to best approximate this formula in multiple imputation:

➡ Use Z, D and $H_0(T)$, and possibly an interaction term as predictor variables.

This often works satisfactorily **if covariate effects and cumulative incidences are rather small**.

Results from the Literature

Problem: in practice $H_0(T)$ is unspecified.

Two main ideas:

- If covariate effects β_x and β_z are small: $H_0(t) \approx H(t)$
  ➡ $H(t)$ can be approximated by the **Nelson-Aalen estimator**.
- **Estimate $H_0(T)$ in an additional step** inside MICE
  ➡ fit a Cox model on the imputed data in each iteration

Results from the Literature

Problem: in practice $H_0(T)$ is unspecified.

Two main ideas:

- If covariate effects β_x and β_z are small: $H_0(t) \approx H(t)$
  ➡ $H(t)$ can be approximated by the **Nelson-Aalen estimator**.
- **Estimate $H_0(T)$ in an additional step** inside MICE
  ➡ fit a Cox model on the imputed data in each iteration

Conclusion (White and Royston 2009):
Use Z, D **and the Nelson-Aalen estimator** $\hat{H}(T)$ as predictors for the imputation of X.

Results from the Literature

Note:

- Neither of these approaches takes into account uncertainty about $H_0(t)$ (but the impact is likely to be small).
- Using the Nelson-Aalen estimator is an approximation
➡ some **bias towards the null** should be expected when covariates have large effects.

Imputation with mice

Example Data:

```
head(survdat)
```

```
##          Time event   x2         x1         x3
## 1 13.156463     0    0  0.7385227 -0.1601367
## 2 12.540724     1 <NA> -0.5147605         NA
## 3  3.344187     1    0 -1.6401813         NA
## 4  9.547701     0    0         NA         NA
## 5  4.077281     0    0         NA  0.1941661
## 6  8.646488     0    0  0.7382478  0.2994167
```

Calculate the Nelson-Aalen estimator using `nelsonaalen()` from the package **mice**:

```
survdat$H0 <- nelsonaalen(survdat, timevar = Time, statusvar = event)
```

Imputation with mice

```
# setup run
imp0 <- mice(survdat, maxit = 0)
meth <- imp0$method
pred <- imp0$predictorMatrix

# specify normal imputation for continuous covariates
meth[c("x1", "x3")] <- "norm"

# remove event time from predictor (high correlation with H0)
pred[, "Time"] <- 0
```

```
pred
```

```
##       Time event x2 x1 x3 H0
## Time     0     1  1  1  1  1
## event    0     0  1  1  1  1
## x2       0     1  0  1  1  1
## x1       0     1  1  0  1  1
## x3       0     1  1  1  0  1
## H0       0     1  1  1  1  0
```

Imputation with mice

```r
# run the imputation
survimp <- mice(survdat, maxit = 10, method = meth,
                predictorMatrix = pred, printFlag = FALSE)
```

To obtain the pooled results, we first fit the model of interest

```r
library("survival")
cox_mice <- with(survimp, coxph(Surv(Time, event) ~ x1 + x2 + x3))
```

and pool and summarize the results.

```r
res_mice_surv <- summary(pool(cox_mice), conf.int = TRUE)
```

Imputation with JointAI

Two options:

- `coxph_imp()`
  proportional hazards model with flexible baseline hazard
- `survreg_imp()`
  parametric (Weibull) model (AFT model)

Imputation with JointAI

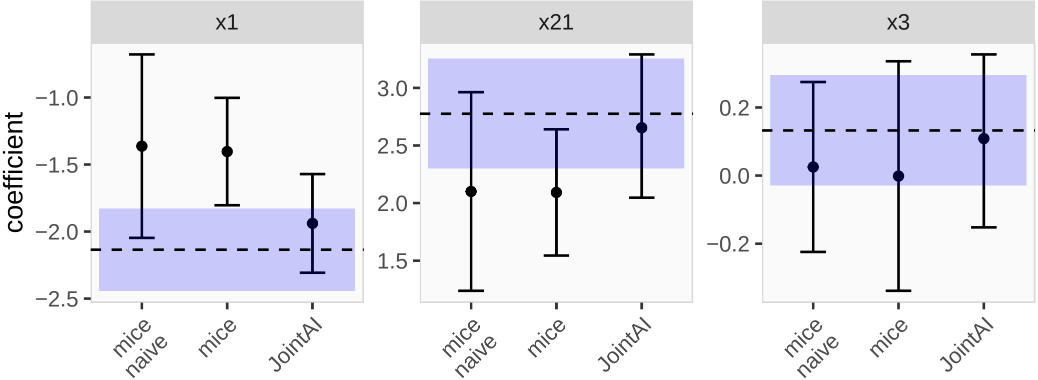
Two options:

- `coxph_imp()`
  proportional hazards model with flexible baseline hazard
- `survreg_imp()`
  parametric (Weibull) model (AFT model)

```
JointAI_cox <- coxph_imp(Surv(Time, event) ~ x1 + x2 + x3, data = survdat,
                         n.iter = 1500)

JointAI_survreg <- survreg_imp(Surv(Time, event) ~ x1 + x2 + x3,
                               data = survdat, n.iter = 1500)
```

Comparison of the Results



Note that the **true effects** (log HR) of $x1$ and $x2$ are **very large** (-2 and 2.5, respectively), and represent the setting where the approximation by the Nelson-Aalen estimate is **expected to be biased**.

References

White, Ian R, and Patrick Royston. 2009. "Imputing Missing Covariate Values for the Cox Model." *Statistics in Medicine* 28 (15): 1982–98.