EP16: Missing Values in Clinical Research: Multiple Imputation

13. Imputation of Survival Data

Nicole S. 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 would be

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

White and Royston (2009) derived versions of this model for different settings and investigated how to best approximate it.

They found that using

- Z, D and $H_0(T)$, and
- possibly an interaction term

as predictor variables may work 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, $H_0(t) \approx H(t)$, which can be approximated by the **Nelson-Aalen estimator**.
- **Estimate $H_0(T)$ in an additional step** inside the MICE procedure by fitting a Cox model on the imputed data.

Neither of these approaches takes into account uncertainty about $H_0(t)$ (but the impact is likely to be 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, $H_0(t) \approx H(t)$, which can be approximated by the **Nelson-Aalen estimator**.
- **Estimate $H_0(T)$ in an additional step** inside the MICE procedure by fitting a Cox model on the imputed data.

Neither of these approaches takes into account uncertainty about $H_0(t)$ (but the impact is likely to be small).

Based on results from their simulation study, White et al. conclude that **using Z, D and the Nelson-Aalen estimator** $\hat{H}(T)$ as predictors for the imputation of X worked best.

However, some **bias towards the null** should be expected when covariates have large effects.

Imputation with mice

In **mice**, `nelsonaalen()` can be used to **calculate the Nelson-Aalen estimator**, to use it as covariate in the imputation.

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

Then, we can prepare the imputation using the same steps as in previous examples:

```
# 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
```

Imputation with mcie

With the modified arguments `method` and `predictorMatrix` we 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

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

and pool and summarize the results.

```
res_mice_surv <- summary(pool(cox_mice, dfcom = 99999), conf.int = TRUE)
```

The warning message refers to the way the degrees of freedom for the formulas we saw in Part I (slide **??**) are calculated and can be ignored.

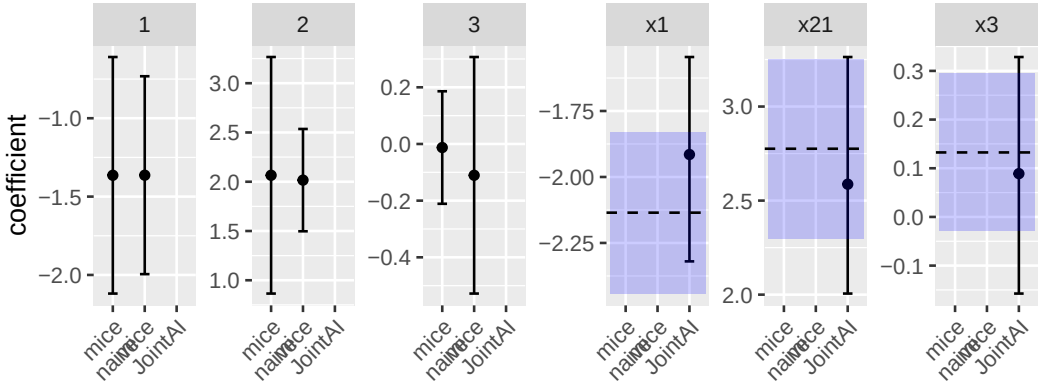
Imputation with JointAI

JointAI has implemented two models for right-censored survival data. The Cox PH model and a parametric Weibull model.

The Cox model is implemented in counting process notation and may take longer to calculate when there are many event times.

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

Comparison of the Results



The naive mice approach, and mice using the Nelson-Aalen estimator give very biased results for the effects of $x1$ and $x2$, but performed acceptably well for $x3$.

Comparison of the Results

Your Turn!

Practical

Imputation of Survival Data **html**

References

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