Solutions: Chapter 11

Exercise 1

The data are in the **kmi** package

```
> require(kmi)
> data(icu.pneu)
```

Coding a new outcome variable (see beginning of 11.2.3)

```
> icu.pneu$outcome <- with(icu.pneu, status * event)
```

Coding of the subdistribution hazard times

```
> icu.pneu$sh.time <- with(icu.pneu,
+                          ifelse(outcome == 3, adm.cens.exit, stop))
```

Proportional subdistribution hazards model

```
> cox.sh.death <- coxph(Surv(start, sh.time, outcome == 2) ~ pneu,
+                       icu.pneu)
```

Exercise 2

```
> data(sir.cont)
```

a)
First, we need to transform `sir.cont` to fit a cox model with a time-dependent covariate

```
> sir.cont.tdc <- sir.cont[order(sir.cont$id, sir.cont$time), ]
> ind <- c(1, diff(sir.cont.tdc$id))
> indd <- c(diff(sir.cont.tdc$id), 1)
> sir.cont.tdc$start <- 0
> sir.cont.tdc$start[ind == 0] <- sir.cont.tdc$time[indd == 0]
> sir.cont.tdc$stop <- sir.cont.tdc$time
> sir.cont.tdc$ventil <- sir.cont.tdc$from
> sir.cont.tdc$status <- as.numeric(sir.cont.tdc$to == 2)
```

Fit of the Cox model

```
> cox.ventil <- coxph(Surv(start, stop, status) ~ ventil, sir.cont.tdc)
```

b)
Creation of the variable `prior.ventil` that indicates whether a patient has been ventilated before

```
> sir.cont.tdc$prior.ventil <- unlist(by(sir.cont.tdc, sir.cont.tdc$id, function(x) {
+     ii <- which(x$ventil == 1)
+     pv <- matrix(0, nrow = nrow(x))
+     if (sum(x$ventil) >= 1) {
+         if (ii[1] < nrow(x)) {
+             pv[(ii[1] + 1):nrow(x)] <- 1
+         }
+     }
+     pv
+ }))
> cox.prior <- coxph(Surv(start, stop, status) ~ ventil + prior.ventil,
+                    sir.cont.tdc)
```

Exercise 3

We take the data `dat1` and `dat2` which were simulated in chapter 10 and add this new covariate names `tdp`

```
> dat1$tdp <- unlist(by(dat1, dat1$id, function(dd) {
+     ind <- which(dd$from == 2)[1]
+     if (!is.na(ind)) {
+         aa <- c(rep(0, ind - 1), rep(1, nrow(dd) - (ind - 1)))
+     } else {
+         aa <- rep(0, nrow(dd))
+     }
+     aa
+ }))
> dat2$tdp <- unlist(by(dat2, dat2$id, function(dd) {
+     ind <- which(dd$from == 2)[1]
+     if (!is.na(ind)) {
+         aa <- c(rep(0, ind - 1), rep(1, nrow(dd) - (ind - 1)))
+     } else {
+         aa <- rep(0, nrow(dd))
+     }
+     aa
+ }))
```

Analysis assuming an illness-death model:

- From state 0 to state 1 for 1st scenario

```
> cox.scenar1.01 <- coxph(Surv(entry, exit, to == 1) ~ Z + tdp + cluster(id),
+                         dat1, subset = from %in% c(0, 2))
```

- From state 0 to state 3 for 1st scenario

```
> cox.scenar1.03 <- coxph(Surv(entry, exit, to == 3) ~ Z + tdp + cluster(id),
+                         dat1, subset = from %in% c(0, 2))
```

- From state 0 to state 1 for 2nd scenario

```
> cox.scenar2.01 <- coxph(Surv(entry, exit, to == 1) ~ Z + tdp + cluster(id),
+                         dat2, subset = from %in% c(0, 2))
```

- From state 0 to state 3 for 2nd scenario

```
> cox.scenar2.03 <- coxph(Surv(entry, exit, to == 3) ~ Z + tdp + cluster(id),
+                         dat2, subset = from %in% c(0, 2))
```

Scenario 1

```
> cox.c.scenar1.01 <- coxph(Surv(entry, exit, to == 1) ~ Z + tdp + cluster(id),
+                           dat1, subset = from == 0)
> cox.c.scenar1.21 <- coxph(Surv(entry, exit, to == 1) ~ Z + tdp + cluster(id),
+                           dat1, subset = from == 2)
> cox.c.scenar1.03 <- coxph(Surv(entry, exit, to == 3) ~ Z + tdp + cluster(id),
+                           dat1, subset = from == 0)
> cox.c.scenar1.23 <- coxph(Surv(entry, exit, to == 3) ~ Z + tdp + cluster(id),
+                           dat1, subset = from == 2)
```

Scenario 2

```
> cox.c.scenar2.01 <- coxph(Surv(entry, exit, to == 1) ~ Z + tdp + cluster(id),
+                           dat1, subset = from == 0)
> cox.c.scenar2.21 <- coxph(Surv(entry, exit, to == 1) ~ Z + tdp + cluster(id),
+                           dat1, subset = from == 2)
> cox.c.scenar2.03 <- coxph(Surv(entry, exit, to == 3) ~ Z + tdp + cluster(id),
+                           dat1, subset = from == 0)
> cox.c.scenar2.23 <- coxph(Surv(entry, exit, to == 3) ~ Z + tdp + cluster(id),
+                           dat1, subset = from == 2)
>
```

Exercise 4

We consider a binary time-dependent exposure. A correct analysis could be performed within the multistate model of Figure 1 (left), whereas the time-dependent bias analysis is represented by Figure 1 (right).

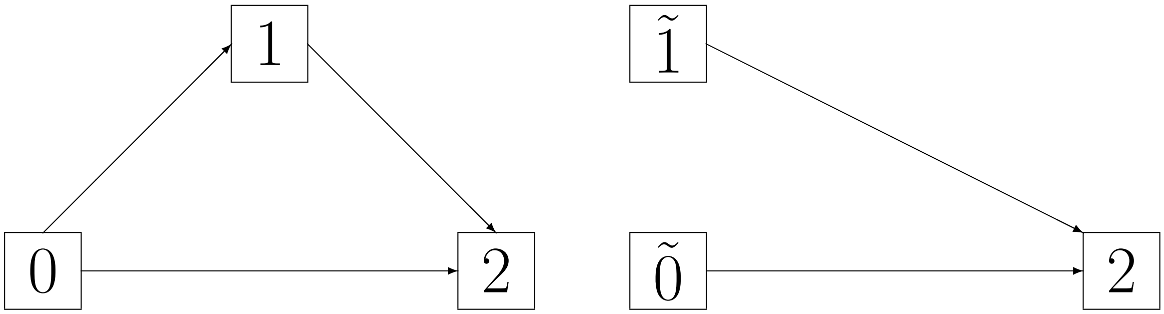

Figure 1: Correct multistate model on the left, with time-dependent bias on the right.

- Show that an analysis subject to time-dependent bias underestimates the cumulative hazard of reaching the absorbing state for exposed individuals.

  That means we have to show that

  $$\sum_{t_k \leq t} \frac{\Delta N_{\tilde{1}2}(t_k)}{Y_{\tilde{1}}(t_k)} < \sum_{t_k \leq t} \frac{\Delta N_{12}(t_k)}{Y_1(t_k)}.$$

  Proof. The number of events at time t are equal in both models, i.e., $\Delta N_{\tilde{1}2}(t) = \Delta N_{12}(t)$. However, $Y_{\tilde{1}}(t) > Y_1(t)$. □

- Show that the biased analysis overestimates the corresponding cumulative hazard for non-exposed individuals. I.e., show that

  $$\sum_{t_k \leq t} \frac{\Delta N_{\tilde{0}2}(t_k)}{Y_{\tilde{0}}(t_k)} > \sum_{t_k \leq t} \frac{\Delta N_{02}(t_k)}{Y_0(t_k)}.$$

  Proof. $\Delta N_{\tilde{0}2}(t) = \Delta N_{02}(t)$ and $Y_{\tilde{0}}(t) < Y_0(t)$. □