Practical Time-varying confounding: ACTG 320 study

Solutions

Kelly Van Lancker

2/13/2024

Installing and Loading R Packages

The following packages and their dependencies need to be installed:

- ipw - Estimating Inverse Probability Weights
- survival - Contains the core survival analysis routines

```
required_packages <-
  c("ipw", "survival")

install.packages(required_packages)
```

Once the required packages are installed, they can be loaded using `library()`

```
library(ipw)
library(survival)
```

Background: ACTG 320 Study Design

The AIDS Clinical Trials Group (ACTG) conducted protocol 320 (ACTG 320), a double-blind study aimed at evaluating two antiretroviral therapy (ART) regimens for AIDS treatment. One regimen, termed "highly-active ART regimen" (HAART), included the addition of Indinavir to the existing regimen of zidovudine, stavudine, and lamivudine (ZDV/D4T/3TC - combination ART regimen). Eligibility for the trial required participants to have a CD4 cell count of no more than 200 cells per cubic millimeter and a minimum of three months of prior zidovudine therapy.

Hammer et al. (1997) documented the results of this study. 1156 patients, who had not undergone treatment with lamivudine or protease inhibitors before, were stratified according to CD4 cell count (50 or fewer vs. 51 to 200 cells per cubic millimeter), and then randomly allocated to one of two daily treatment regimens: either 600 mg of zidovudine and 300 mg of lamivudine, or the same regimen along with 2400 mg of indinavir. The primary endpoint was the duration time to the onset of acquired immunodeficiency syndrome (AIDS) or death. The proportion of patients whose disease progressed to AIDS or death was lower with indinavir, zidovudine (or stavudine), and lamivudine than with zidovudine (or stavudine) and lamivudine alone (estimated hazard ratio, 0.50; 95 percent confidence interval, 0.33 to 0.76; $P = 0.001$).

As discussed in Cain & Cole (2009), around a quarter of patients dropped out or stopped adhering to their assigned therapy for reasons other than toxicity. Further, the time to non-adherence differed between study

arms. As a result, there were concerns that the treatment-policy analysis estimates would be shrunken towards the null. In this practical, our aim is to replicate the analysis in their paper and address the potential differences in compliance in the two arms in the subsequent analyses. We hereby consider the following hypothetical estimand: what would the hazard ratio be if everyone had stayed on the treatment they were assigned to.

Loading the data and data management

Starting from original dataset (this part can be skipped)

We can download the orginal dataset from https://eclass.uoa.gr/modules/document/index.php?course=MED1174&openDir=/5aabb509gI5h/5abb8578jC7H, and then load it into R as follows:

```
actg320_original <- read.csv("actg320_original.csv", header=TRUE, stringsAsFactors=FALSE)
```

However, as our goal is to replicate the analysis in Cain and Cole (2009), we still need to do some data management to define the **non-compliance** and **event (AIDS or death)** variable. For the data management we follow: https://eclass.uoa.gr/modules/document/file.php/MED1174/R%20Labs/A%20case%20study%3A%20The%20ACTG%20320%20protocol/actg320.pdf.

```
# Data management

# Fix CD4 counts
actg320_original$cd4.sqrt=sqrt(actg320_original$CD410*10)
actg320_original$bcd4.sqrt=sqrt(actg320_original$BCD410*10)

# Create first and last observation
actg320_original$first<-!duplicated(actg320_original$pidnum)
actg320_original$last<-!duplicated(actg320_original$pidnum, fromLast = TRUE)

# Allow event to be equal to 1 only in the final observation
actg320_original$aidsdeath<-(actg320_original$last==1)*(actg320_original$event==1)

# Allow dropout event to be equal to 1 only in the final observation
actg320_original$ltfu<-(actg320_original$last==1)*(actg320_original$event==2)

# Create non-compliance endpoint (at minimum between toxicity and non-compliance)
actg320_original$non.compliance<-(actg320_original$comply==0 & actg320_original$toxic==0)
actg320_original$non.compliance<-ifelse((actg320_original$non.compliance==0
                                & actg320_original$last==1 & actg320_original$event==2),1,
                               actg320_original$non.compliance)

# Additional cases
actg320_original$non.compliance<-ifelse((actg320_original$non.compliance==0
& actg320_original$last==1 & actg320_original$event==2),1,
actg320_original$non.compliance)

# Start and stop day of each interval
actg320_original$tstart = actg320_original$day
actg320_original$tstop = actg320_original$day+1

actg320 = actg320_original
```

Dataset after data management

Here, we immediately load the dataset that can be obtained after doing the data management above on the original dataset. We input the longitudinal ACTG 320 data as follows:

```
actg320 <- read.csv("actg320.csv", header=TRUE, stringsAsFactors=FALSE)
```

Exploring the data

In this part we will familiarize ourselves with the data.

Important variables (for our analysis) in this dataset:

- 'pidnum': patient identification number.
- 'rnd': treatment indicator; 1 if treatment includes indinavir, 0 otherwise.
- 'non.compliance': 1 if the patients is no longer complying to the assigned treatment, 0 otherwise.
- 'aidsdeath': event of interest; 1 if death or incurred AIDS, 0 otherwise.
- 'male': 1 if male, 0 if female.
- 'IDU': indinavir drug use history; 1 if yes, 0 if no.
- 'black': 1 if black, 0 otherwise.
- 'hisp': 1 if hispanic, 0 otherwise.
- 'AZT1yr': greater than one year Zidovudine use at randomization; 1 if yes, 0 if no.
- 'age10': age at randomization (by 10-year increments).
- 'cd4.sqrt':time-varying CD4 count.
- 'bcd4.sqrt': baseline CD4 count.

```
head(actg320)
```

```
##   pidnum male black hisp age10 rnd BCD410 AZT1yr IDU event day CD410 pain
## 1  10333    1     0    0     4   0    9.5      1   0     2   0   9.5    0
## 2  10333    1     0    0     4   0    9.5      1   0     2   1  12.0    0
## 3  10333    1     0    0     4   0    9.5      1   0     2   2  12.0    0
## 4  10333    1     0    0     4   0    9.5      1   0     2   3  12.0    0
## 5  10333    1     0    0     4   0    9.5      1   0     2   4  12.0    0
## 6  10333    1     0    0     4   0    9.5      1   0     2   5  12.0    0
##   fatigue headache fever comply toxic version  cd4.sqrt bcd4.sqrt first  last
## 1       0        0     0      1     0     V01  9.746794  9.746794  TRUE FALSE
## 2       0        0     0      1     0     V01 10.954451  9.746794 FALSE FALSE
## 3       0        0     0      1     0     V01 10.954451  9.746794 FALSE FALSE
## 4       0        0     0      1     0     V01 10.954451  9.746794 FALSE FALSE
## 5       0        0     0      1     0     V01 10.954451  9.746794 FALSE FALSE
## 6       0        0     0      1     0     V01 10.954451  9.746794 FALSE FALSE
##   aidsdeath ltfu non.compliance tstart tstop
## 1         0    0              0      0     1
## 2         0    0              0      1     2
```

```
## 3         0    0              0      2     3
## 4         0    0              0      3     4
## 5         0    0              0      4     5
## 6         0    0              0      5     6
```

We first consider the data for the patient with `pidnum==10350`.

```
# Data for patient 10350
View(actg320[actg320$pidnum==10350,
             c("pidnum", "rnd", "tstart", "tstop", "non.compliance", "aidsdeath")])
```

We observe that this patient was randomized to the treatment including indinavir, and remained compliant till he left the study.

We now consider the data for the patient with `pidnum==10333`.

```
# Data for patient 10333
View(actg320[actg320$pidnum==10333,
             c("pidnum", "rnd", "tstart", "tstop", "non.compliance", "aidsdeath")])
```

This patient was randomized to the standard therapy, remained compliant until day 28, at which point he stopped complying, and continued in the study till day 132.

We now consider the data for the patient with `pidnum==12566`.

```
# Data for patient 12566
View(actg320[actg320$pidnum==12566,
             c("pidnum", "rnd", "tstart", "tstop", "non.compliance", "aidsdeath")])
```

This patient was randomized to the standard therapy, remained compliant and died on day 52.

Finally, we consider the data for the patient with `pidnum==10719`.

```
# Data for patient 10719
View(actg320[actg320$pidnum==10719,
             c("pidnum", "rnd", "tstart", "tstop", "non.compliance", "aidsdeath")])
```

This patient was randomized to the standard therapy, remained compliant until day 77, at which point he stopped complying, and continued in the study till day 113, at which he died.

Intention-to-treat analysis

Question 1.

Our first aim is to replicate the results of Hammer et al. (1997). Conduct an unadjusted analysis using the function `coxph`, and report the corresponding hazard ratio with p-value.

```
coxph(Surv(tstart, tstop, aidsdeath) ~ rnd, data = actg320, cluster = pidnum)
```

```
## Call:
## coxph(formula = Surv(tstart, tstop, aidsdeath) ~ rnd, data = actg320,
```

```
##     cluster = pidnum)
##
##        coef exp(coef) se(coef) robust se      z       p
## rnd -0.6815    0.5059   0.2149    0.2154 -3.164 0.00155
##
## Likelihood ratio test=10.6  on 1 df, p=0.001128
## n= 287133, number of events= 96
```

IPW implementation: unstabilized weights

Question 2.

For each of the four patients above, assuming non-compliance happens at the end of the time interval, from which time interval onwards would you censor them?

Question 3.

We now will calculate the unstabilized weights for IPW using a Cox model. We will do this via the function `ipwtm` (using `type = "cens"`) in the `ipw` package. The propensity score/weights should be based on treatment `rnd`, the baseline variables `bcd4.sqrt`, `male`, `IDU`, `black`, `hisp`, `AZT1yr` and `age10`, and the time-varying confounder `cd4.sqrt`. Save the `ipwtm` object in `temp2` and use the code below to include the weights and selection variable in the ACTG 320 dataset.

```
# Calculate weights
temp2 <- ipwtm(
exposure = non.compliance,
family = "survival",
denominator = ~ rnd + bcd4.sqrt + male + IDU + black + hisp + AZT1yr + age10 +
cd4.sqrt,
id = pidnum,
tstart = tstart,
timevar = tstop,
type = "cens",
data = actg320)

# Include weights and selection variable in actg 320 dataset
actg320$selvar<-temp2$selvar
actg320$ipw.weights<-temp2$ipw.weights
```

Question 4.

Investigate the obtained dataset, especially the inverse probability weights and the variable `selvar`, which becomes 0 at the first interval after the first occurence of non-compliance. What happens with the weights (and `selvar`) when `non.compliance` becomes 1?

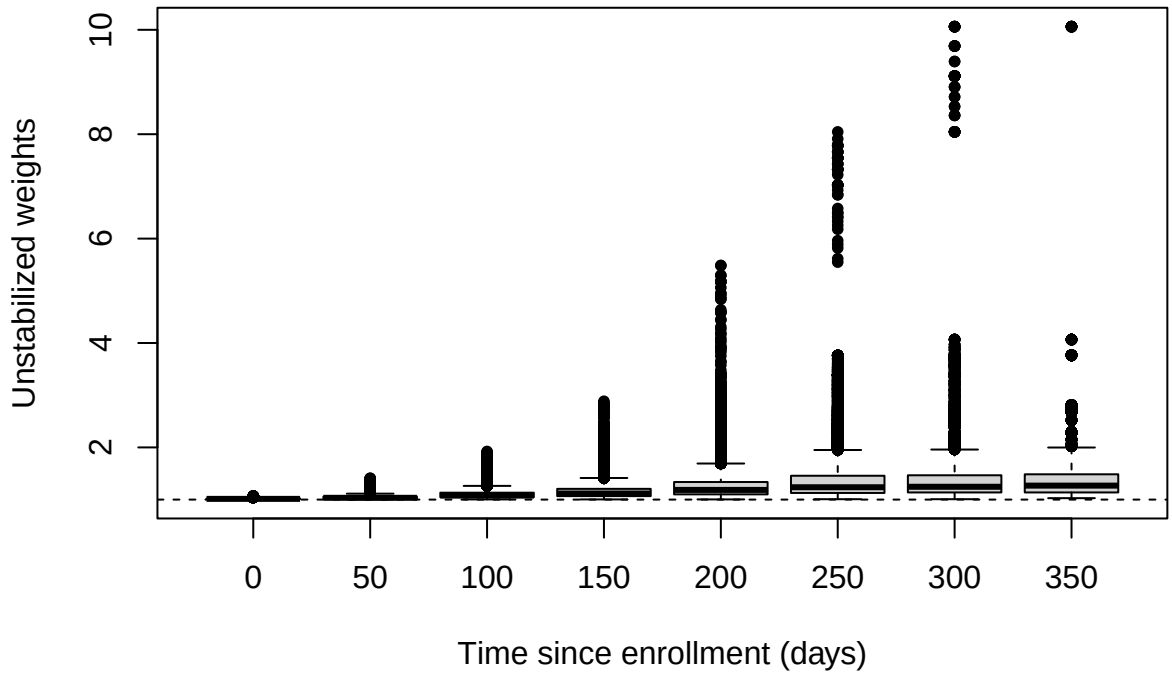
```
# Data for patient 10350
View(actg320[actg320$pidnum==10350,
             c("pidnum", "rnd", "tstart", "tstop", "non.compliance",
               "aidsdeath", "ipw.weights", "selvar")])

# Data for patient 10333
```

```
View(actg320[actg320$pidnum==10333,
             c("pidnum", "rnd", "tstart", "tstop", "non.compliance",
               "aidsdeath", "ipw.weights", "selvar")])

# Data for patient 12566
View(actg320[actg320$pidnum==12566,
             c("pidnum", "rnd", "tstart", "tstop", "non.compliance",
               "aidsdeath", "ipw.weights", "selvar")])

# Data for patient 10719
View(actg320[actg320$pidnum==10719,
             c("pidnum", "rnd", "tstart", "tstop", "non.compliance",
               "aidsdeath", "ipw.weights", "selvar")])
```

Question 5.

Plot the unstabilized weights using the function `ipwplot`.

```
ipwplot(actg320$ipw.weights, timevar = actg320$tstop,
        binwidth = 50,
        xlab = "Time since enrollment (days)",
        ylab = "Unstabilized weights",
        logscale = F)
```



Question 6.

Perform the weighted Cox proportional hazards analysis for the event of interest (death or incurrence of aids), using robust standard errors. Note that the considered dataset only includes the time intervals with `actg320$selvar==1`: `data=actg320[actg320$selvar==1,]`.

```
coxph(Surv(tstart, tstop, (aidsdeath)) ~ rnd + cluster(pidnum),
data = actg320[actg320$selvar==1,], weights=ipw.weights)
```

```
## Call:
## coxph(formula = Surv(tstart, tstop, (aidsdeath)) ~ rnd, data = actg320[actg320$selvar ==
##     1, ], weights = ipw.weights, cluster = pidnum)
##
##        coef exp(coef) se(coef) robust se      z      p
## rnd -0.7831    0.4570   0.2358    0.2580 -3.035 0.0024
##
## Likelihood ratio test=11.88  on 1 df, p=0.0005686
## n= 253438, number of events= 68
```

What is your conclusion? Is it in line with the analysis in Hammer et al. (1997), which didn't account for non-compliance?

IPW implementation: stabilized weights

Question 7.

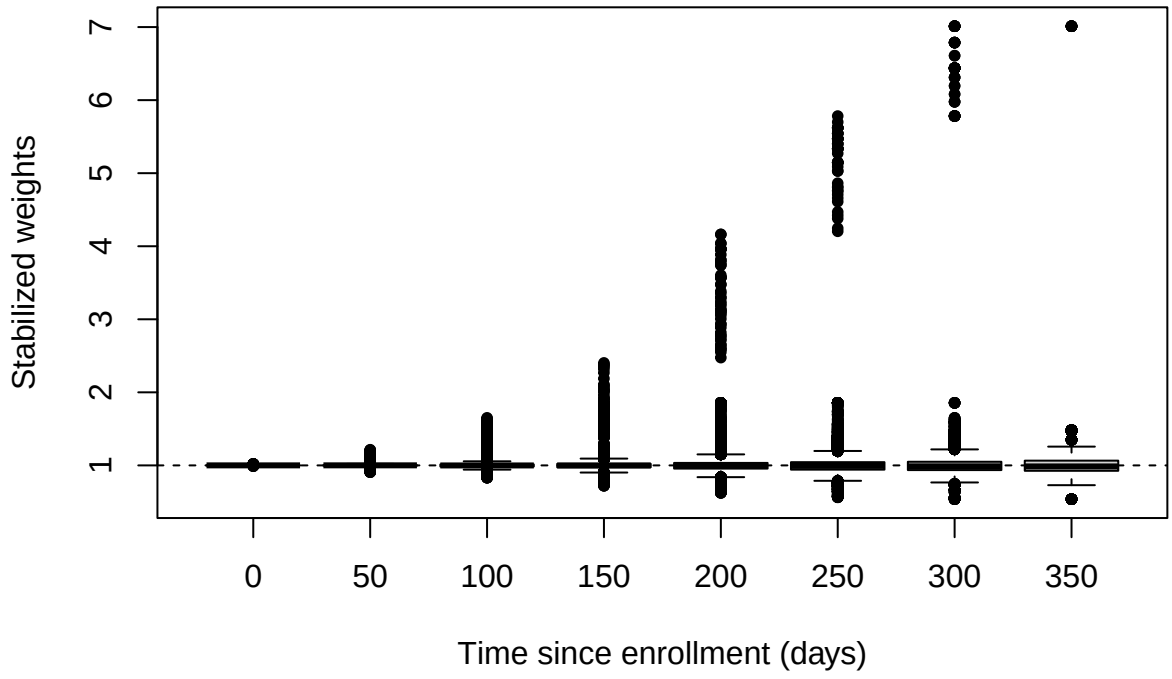
We now will calculate the stabilized weights for IPW using a Cox model. We will do this via the function `ipwtm` (using `type = "cens"`) in the `ipw` package. The propensity score/weights should be based on treatment `rnd`, the baseline variables `bcd4.sqrt`, `male`, `IDU`, `black`, `hisp`, `AZT1yr` and `age10`, and the time-varying confounder `cd4.sqrt`. Save the `ipwtm` object in `temp2` and use the code below to include the weights and selection variable in the ACTG 320 dataset.

```
# Calculate weights
temp2 <- ipwtm(
exposure = non.compliance,
family = "survival",
numerator = ~ rnd + bcd4.sqrt + male + IDU + black + hisp + AZT1yr +age10,
denominator = ~ rnd + bcd4.sqrt + male + IDU + black + hisp + AZT1yr + age10 +
cd4.sqrt,
id = pidnum,
tstart=tstart,
timevar = tstop,
type = "cens",
data = actg320)

# Include weights and selection variable in actg 320 dataset
actg320$selvar<-temp2$selvar
actg320$ipw.weights<-temp2$ipw.weights
```

Question 8.

Plot the stabilized weights using the function `ipwplot`.

```
ipwplot(actg320$ipw.weights, timevar = actg320$tstop,
        binwidth = 50,
        xlab = "Time since enrollment (days)",
        ylab = "Stabilized weights",
        logscale = F)
```



Compare the distribution of stabilized and unstabilized weights. What do you observe?

Question 9.

Perform the weighted Cox proportional hazards analysis for the event of interest (death or incurrence of aids), using robust standard errors. Note that the considered dataset only includes the time intervals with `actg320$selvar==1`: `data=actg320[actg320$selvar==1,]`. Note that, as stabilized weights reduce to one if there is only confounding by baseline covariates, adjustment for baseline covariates in the Cox model is crucial now. What is your conclusion? Is it in line with the analysis with unstabilized weights and/or the analysis in Hammer et al. (1997)?

```
coxph(Surv(tstart, tstop, (aidsdeath)) ~ rnd + bcd4.sqrt + male + IDU + black
      + hisp + AZT1yr + age10 + cluster(pidnum),
data = actg320[actg320$selvar==1,],
weights=ipw.weights)
```

```
## Call:
## coxph(formula = Surv(tstart, tstop, (aidsdeath)) ~ rnd + bcd4.sqrt +
##     male + IDU + black + hisp + AZT1yr + age10, data = actg320[actg320$selvar ==
##     1, ], weights = ipw.weights, cluster = pidnum)
##
##                coef exp(coef)  se(coef) robust se      z        p
## rnd       -0.769400  0.463291  0.252866  0.258329 -2.978   0.0029
## bcd4.sqrt -0.226435  0.797371  0.034906  0.031133 -7.273 3.51e-13
## male       0.005117  1.005130  0.363267  0.366521  0.014   0.9889
## IDU       -0.673891  0.509721  0.432800  0.438215 -1.538   0.1241
## black     -0.592598  0.552889  0.317567  0.336584 -1.761   0.0783
## hisp       0.090329  1.094534  0.309943  0.318758  0.283   0.7769
## AZT1yr     0.412655  1.510823  0.281565  0.293426  1.406   0.1596
## age10      0.305258  1.356975  0.132969  0.137793  2.215   0.0267
##
## Likelihood ratio test=70.35  on 8 df, p=4.18e-12
## n= 253438, number of events= 68
```

What is your conclusion? Is it in line with the analysis with unstabilized weights and/or the analysis in Hammer et al. (1997)?

References

Cain, L. E., & Cole, S. R. (2009). Inverse probability-of-censoring weights for the correction of time-varying noncompliance in the effect of randomized highly active antiretroviral therapy on incident AIDS or death. Statistics in medicine, 28(12), 1725-1738.

Hammer, S. M., Squires, K. E., Hughes, M. D., Grimes, J. M., Demeter, L. M., Currier, J. S., ... & Cook, J. C. (1997). A controlled trial of two nucleoside analogues plus indinavir in persons with human immunodeficiency virus infection and CD4 cell counts of 200 per cubic millimeter or less. New England Journal of Medicine, 337(11), 725-733.