Multistate models and recurrent event models

Patrick Breheny

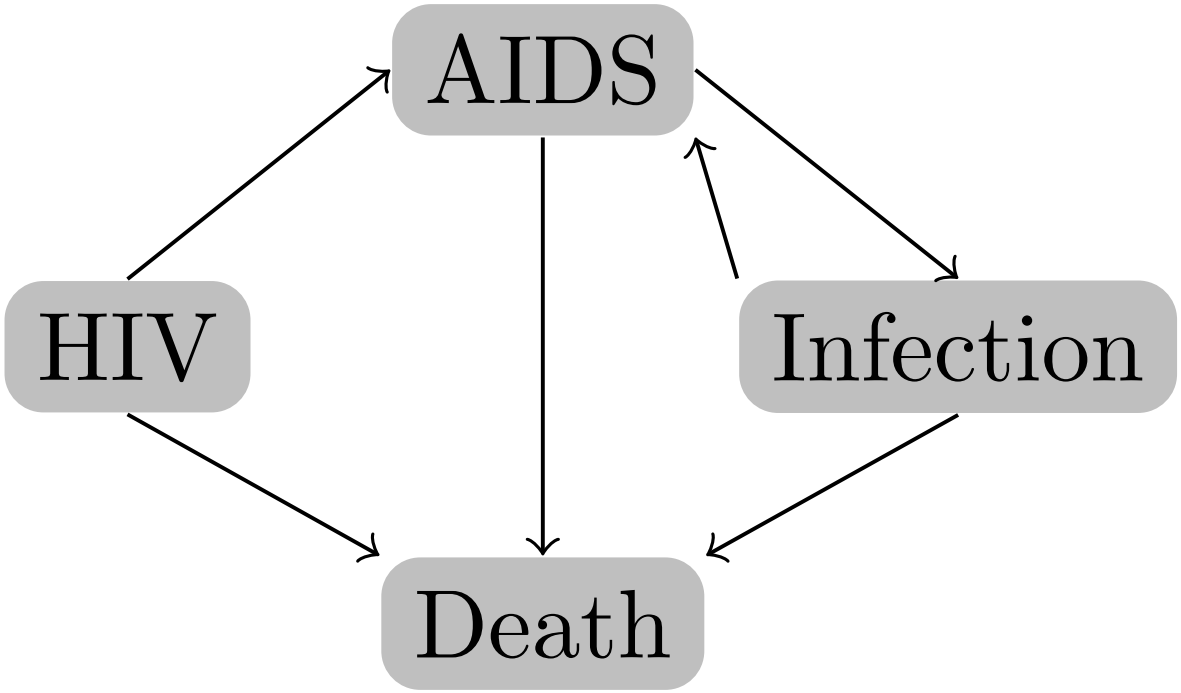
December 7

Introduction

- In this final lecture, we will briefly look at two other kinds of time-to-event data and how the models we've discussed previously can be extended to analyze them
- First, we'll consider multi-state models, which we briefly introduced last time
- The main idea is that as a subject moves through time, they can transition between multiple states, with $S(t)$ denoting their state at time t; our previous setup can be thought of a special case with just two states, alive and dead, with no possibility of the transition dead $\rightarrow$ alive

An illustration

For example, a model for the transitions between various stages of the progression of AIDS might look like:



where infection denotes an opportunistic infection associated with immune deficiency

Transition rates

- Multistate data are modeled according to the *transition rate* $\lambda_{ij}(t)$, which describes the probability of transitioning to state j, given that an individual is in state i at time t:

$$\lambda_{ij}(t) = \lim_{h \to 0} h^{-1} \mathbb{P}\{S(t+h) = j | S(t) = i\}$$

- Implicit in this definition is the idea that $S(t)$ is a *Markov process*, meaning that transition probabilities depend only on the current state $S(t)$ and not the specific path taken to arrive at $S(t)$
- This is equivalent to our definition of the type-specific hazard from the previous lecture; at each state, we have competing risks corresponding to the probabilities of transitioning to the other states of the model

Applying regression methods

- Thus, we could use any of the models we have discussed so far to model transition rates
- For example, a parametric model such as Weibull regression would allow us to estimate transition probabilities continuously, while the semiparametric Cox model would restrict transition probabilities to occur only at times where we have already seen an $i \to j$ transiton
- Note that this would require separate models for each $i \to j$ transition

Homogenous, nonhomogeneous, and renewal models

- The simplest model, of course, is an exponential model, where $\lambda_{ij}(t) = \lambda_{ij}$; because these transition probabilities do not depend on t, this model is said to be *homogeneous*
- If the transition probabilities do depend on time, we have a choice to make:
  - Modeling the transition rates as a function of t, the total time on study (this would be a *nonhomogeneous Markov model*)
  - Modeling the transition rates as a function of the time since arriving at the current state (this would be a *Markov renewal model*)

Introduction

- Another type of time-to-event data that can arise is the possibility that the event can occur multiple times
- Some examples include:
  - Recurrence of cancer
  - Infections
  - Hospital readmissions
  - Relapses for drug abuse
  - Service/repair calls for a machine

Intensity process

- Let $T_1, T_2, \ldots$ denote the time until the first event, second event, and so on for a given subject, and let C denote the total follow-up time
- Note that for recurrent events, everyone is eventually censored; if events can continue to occur, we are never finished observing a subject
- Let $N(t)$ denote the number of events that an individual experiences by time t; the *intensity process* (which may be extended to depend on covariates, of course) is

$$\lambda(t) = \lim_{h \to 0} h^{-1} \mathbb{P}\{N(t+h) - N(t) = 1\}$$

Connection with Cox regression

- If $N(t)$ is a strictly continuous process, one may use the regression methods we have discussed thus far to model it
- This works very similarly to the idea of subject duplication that we discussed with regard to time-dependent covariates: a subject that experiences recurrent events t_1 and t_2, then is censored at c, with $t_1 < t_2 < c$ would be represented with multiple entries in a data frame as

Start	Stop	Event	Recurrence
0	t_1	1	1
t_1	t_2	1	2
t_2	c	0	3

Conditioning on previous recurrences

- A key decision when modeling recurrent events is whether and how to condition on previous recurrences
- The simplest approach would be to assume that events are completely independent, and that the risk of an event at time t is the same regardless of whether it's the first, second, or third recurrence
- Alternatively, one might consider the number of events as a (time-dependent) covariate in the model

Stratification

- It is important to note, however, that using the number of recurrent events experienced so far as a time-dependent covariate assumes proportional hazards across the different recurrences
- This is probably an unrealistic assumption; as an alternative, we might consider allowing each recurrent failure to have its own baseline distribution
- Recall that this can be accomplished through stratification (this time, on a time-dependent covariate) in the Cox model

Gap time models

- Finally, we also face a similar decision that we saw in multi-state modeling: whether to model time to recurrence since the start of the study, or time since the last recurrence
- The latter type of model is known as a *gap time* model; to fit it, we would simply need to reorganize the data as (to revisit our earlier example):

Time	Event	Recurrence
t_1	1	1
$t_2 - t_1$	1	2
$c - t_2$	0	3

- Like origin-time models, we would again have to decide how to condition on previous recurrences (ignore, assume proportional hazards, stratify)

Bladder cancer data

- To briefly illustrate these models, we will analyze data from a VA study of bladder cancer recurrence (`bladder2` in the `survival` package)
- In the study, all patients had bladder tumors when they entered the trial
- At the start of the trial ($t = 0$), these tumors were removed and the patients randomly assigned to receive the anticancer drug thiotepa or a placebo
- In addition, we have covariate data on the initial number of tumors and the size of the largest initial tumor

Origin-time, unconditional model

```
> fit <- coxph(Surv(start, stop, event) ~
                rx + number + size, bladder2)
> summary(fit)
  n= 178, number of events= 112

            coef exp(coef) se(coef)      z Pr(>|z|)
rx      -0.46469   0.62833  0.19973 -2.327 0.019989
number   0.17496   1.19120  0.04707  3.717 0.000202
size    -0.04366   0.95728  0.06905 -0.632 0.527196

Concordance= 0.634  (se = 0.03 )
Rsquare= 0.094   (max possible= 0.994 )
Likelihood ratio test= 17.52  on 3 df,   p=0.0005531
```

Origin-time, # recurrence as covariate

```
> fit <- coxph(Surv(start, stop, event) ~
                rx + number + size + factor(enum), bladder2)
> summary(fit)
  n= 178, number of events= 112

                    coef exp(coef)  se(coef)      z Pr(>|z|)
rx              -0.279944  0.755826  0.205765 -1.361 0.173671
number           0.140337  1.150661  0.051418  2.729 0.006347
size            -0.003751  0.996256  0.070320 -0.053 0.957464
factor(enum)2    0.589260  1.802654  0.256782  2.295 0.021745
factor(enum)3    1.680455  5.367995  0.302358  5.558 2.73e-08
factor(enum)4    1.337645  3.810061  0.351012  3.811 0.000139

Concordance= 0.68  (se = 0.03 )
Rsquare= 0.247   (max possible= 0.994 )
Likelihood ratio test= 50.54  on 6 df,   p=3.662e-09
```

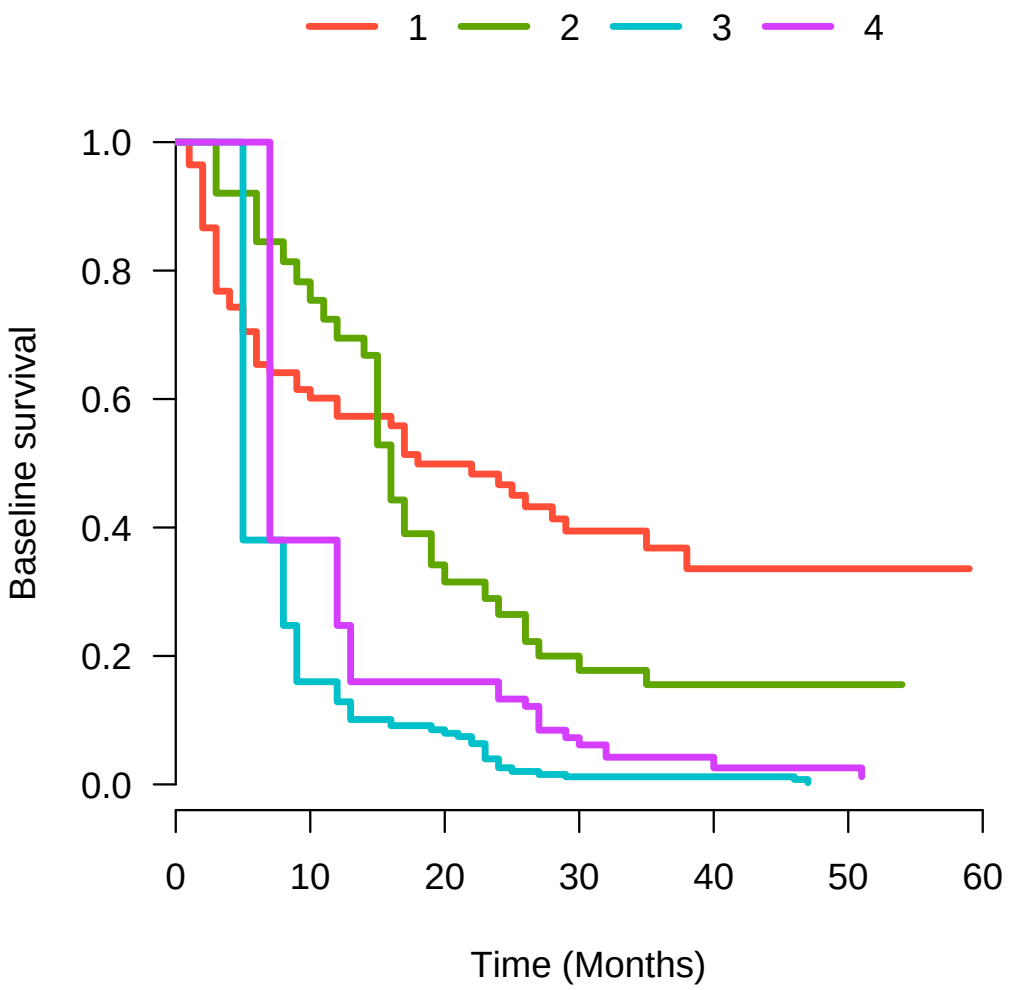
Origin-time, stratified

```
fit <- coxph(Surv(start, stop, event) ~
             rx + number + size + strata(enum), bladder2)
```

Coefficients and Wald tests for treatment:

Recurrence	$\widehat{\beta}$	z	p
1	-0.53	-1.67	0.10
2	-0.50	-1.24	0.21
3	0.14	0.21	0.83
4	0.05	0.06	0.95
Overall	-0.43	-1.96	0.05

Baseline hazards for stratified origin-time model



Gap time, unconditional model

```
> fit <- coxph(Surv(gaptime, event) ~
               rx + number + size, bladder2)
> summary(fit)
  n= 178, number of events= 112

            coef exp(coef) se(coef)      z Pr(>|z|)
rx      -0.37446   0.68766  0.20237 -1.850  0.06426
number   0.15877   1.17207  0.04881  3.253  0.00114
size    -0.02014   0.98006  0.06793 -0.296  0.76686

Concordance= 0.6  (se = 0.032 )
Rsquare= 0.066   (max possible= 0.997 )
Likelihood ratio test= 12.08  on 3 df,   p=0.007119
```

Gap time, # recurrence as covariate

```
> fit <- coxph(Surv(gaptime, event) ~
                rx + number + size + factor(enum), bladder2)
> summary(fit)
  n= 178, number of events= 112

                    coef exp(coef)  se(coef)      z Pr(>|z|)
rx              -0.298270  0.742101  0.205890 -1.449  0.14743
number           0.152642  1.164908  0.051659  2.955  0.00313
size             0.005046  1.005059  0.069182  0.073  0.94186
factor(enum)2    0.181182  1.198633  0.243627  0.744  0.45707
factor(enum)3    0.879807  2.410433  0.269588  3.264  0.00110
factor(enum)4    0.745181  2.106823  0.314762  2.367  0.01791

Concordance= 0.619  (se = 0.032 )
Rsquare= 0.128   (max possible= 0.997 )
Likelihood ratio test= 24.36  on 6 df,   p=0.0004477
```

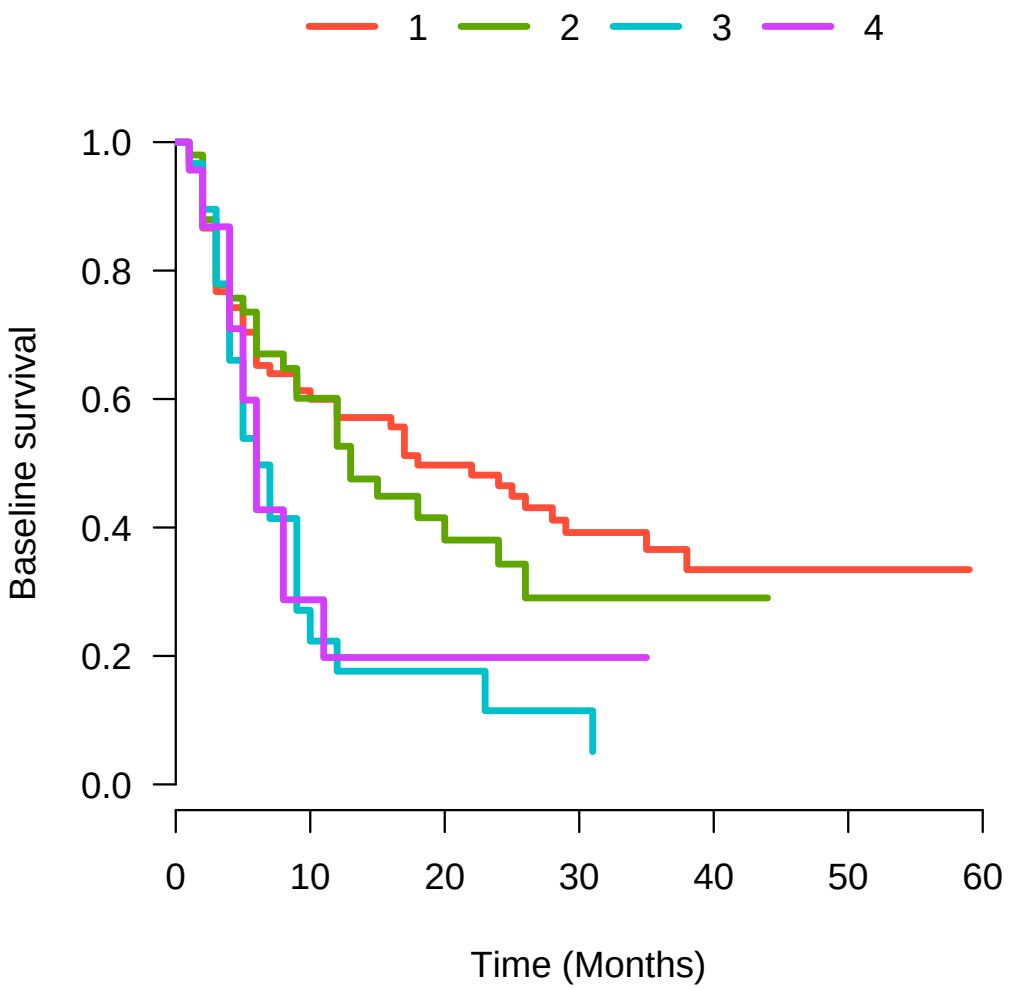
Gap time, stratified

```
fit <- coxph(Surv(gaptime, event) ~
             rx + number + size + strata(enum), bladder2)
```

Coefficients and Wald tests for treatment:

Recurrence	$\widehat{\beta}$	z	p
1	-0.53	-1.67	0.10
2	-0.27	-0.67	0.50
3	0.21	0.38	0.70
4	-0.22	-0.34	0.73
Overall	-0.28	-1.35	0.18

Baseline hazards for stratified gap-time model



Remarks

- Overall, I would tend to place the most trust in the stratified gap-time model in this example
- The general conclusion would be that there seems to be marginal evidence that the treatment is effective at preventing the first recurrence of bladder cancer, but no evidence that the treatment is effect at preventing future recurrences