Discrete-time Survival Analysis using Latent Variables Part 1

Presented by
Katherine E. Masyn, Ph.D.*
kmasyn@ucla.edu

ED231E, UCLA

May 10 & 12, 2004

*Postdoctoral fellow with Johns Hopkins SPH, Department of Mental Health
Supported by NIMH Grant T32-MH018834

Time-to-event data

A record of *when* an event occurs (relative to some “beginning”) for each individual in a sample, e.g., time of death, grade of school drop-out, age of first alcohol use in school-aged children, etc.

RIA Example

Data from a study out of the Research Institute on Addictions at SUNY Buffalo (Bill Fals-Stewart, P.I.) on the drinking and domestic violence behavior of alcohol-dependent men following one of three alcohol abuse treatment regimes.

Data

- 170 men
- Married or cohabitating
- Participated in one of three alcohol treatment programs
- All report at least one episode of domestic violence during the three-month pre-treatment period
- One-year follow-up period discretized into six two-month observation periods

UTEC Example

Preliminary data from a study by the UCLA Urban Teacher Education Collaborative funded by the Stuart Foundation and supervised by Karen Hunter Quartz. The design is a 7-year prospective longitudinal study of the graduates from UCLA's Center X TEP.

Data

- First 6 cohorts of UCLA's Center X TEP graduates: n = 513
- Information on professional status from teaching years 1 to 6 determined by yearly survey with 85-95% response rates
- All are full-time classroom teachers in Year 1.

Time-scales

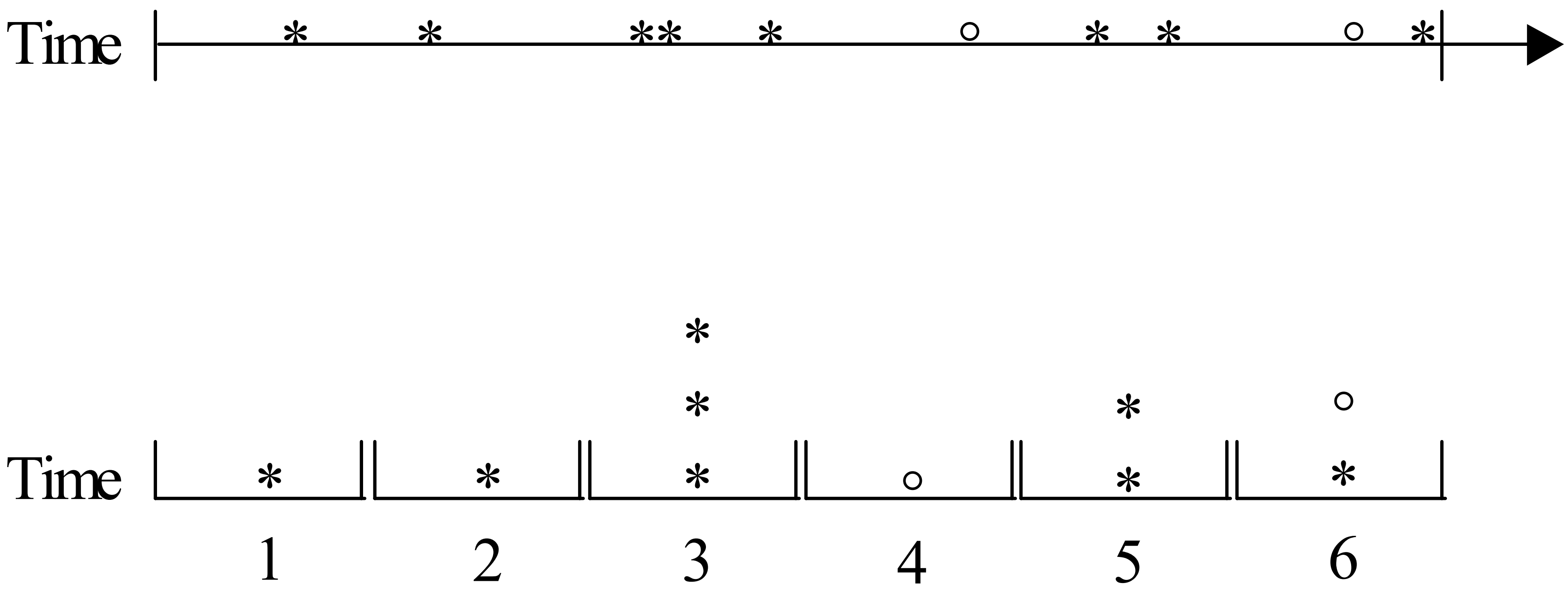
- ***Continuous***

  The “exact” time of an event for each subject is known, e.g., time of death

- ***Discrete***

  1) The timing of an event is continuous but is only recorded for an *interval* of time, e.g., grade of school drop-out.

  2) The timing of an event is itself discrete, e.g., grade retention.


Time
Time
1
2
3
4
5
6

Missing data

- Various mechanisms for missing data in the survival context are referred to under the unifying term, *censoring*, indicating that the event times for some subjects are unknown to the researcher.
- Censoring is usually assumed to be *noninformative* which means that the distribution of censoring times is independent of event times, conditional on the set of observed covariates. (Think: MAR)

Right-censoring

- The most typical survival data is right-censored and this type of *missingness* is the easiest to deal with in the data analysis.

- Right-censoring occurs when a subject in the sample has *not* experienced the event of interest at the end of the observation period. It is assumed that the event eventually occurs sometime after the end of the study.

Survival data

- Assume for now only noninformative, right censoring with no truncation.
- Let C_i be the right-censoring time and T_i be the event time (interval) for individual *i*.
- T_i is observed if $T_i \leq C_i$ and C_i is observed if $T_i > C_i$.
- Let the observed data consist of $\{A_i, \delta_i\}$ where $A_i = \min(T_i, C_i)$ and $\delta_{\mathrm{i}} = \mathrm{I}(T_i \leq C_i)$.

The entire span of observations on a single subject can be summarized by those two numbers, A_i and δ_i, that indicate:

1) the last time period during which the individual was observed, and

2) whether the observation of that individual was discontinued because he/she experienced the event or because he/she was “censored”.

Survival probability

Let *T* be the time interval of the event where T∈{1,2,…,J}

S(j), called the ***survival probability***, is defined as the probability of "surviving" *beyond* time interval *j*, i.e., the probability that the event occurs after interval *j*: $S(j) = P(T > j)$

Hazard probability

h(j), called the ***hazard probability***, is defined as the probability of the event occurring in the time interval *j*, provided it has not occurred prior to *j*:

$$h(j) = P(T = j \mid T \geq j).$$

Essentially, *h(j)* is the probability of the event occurring in time interval *j* among those at-risk in *j*.

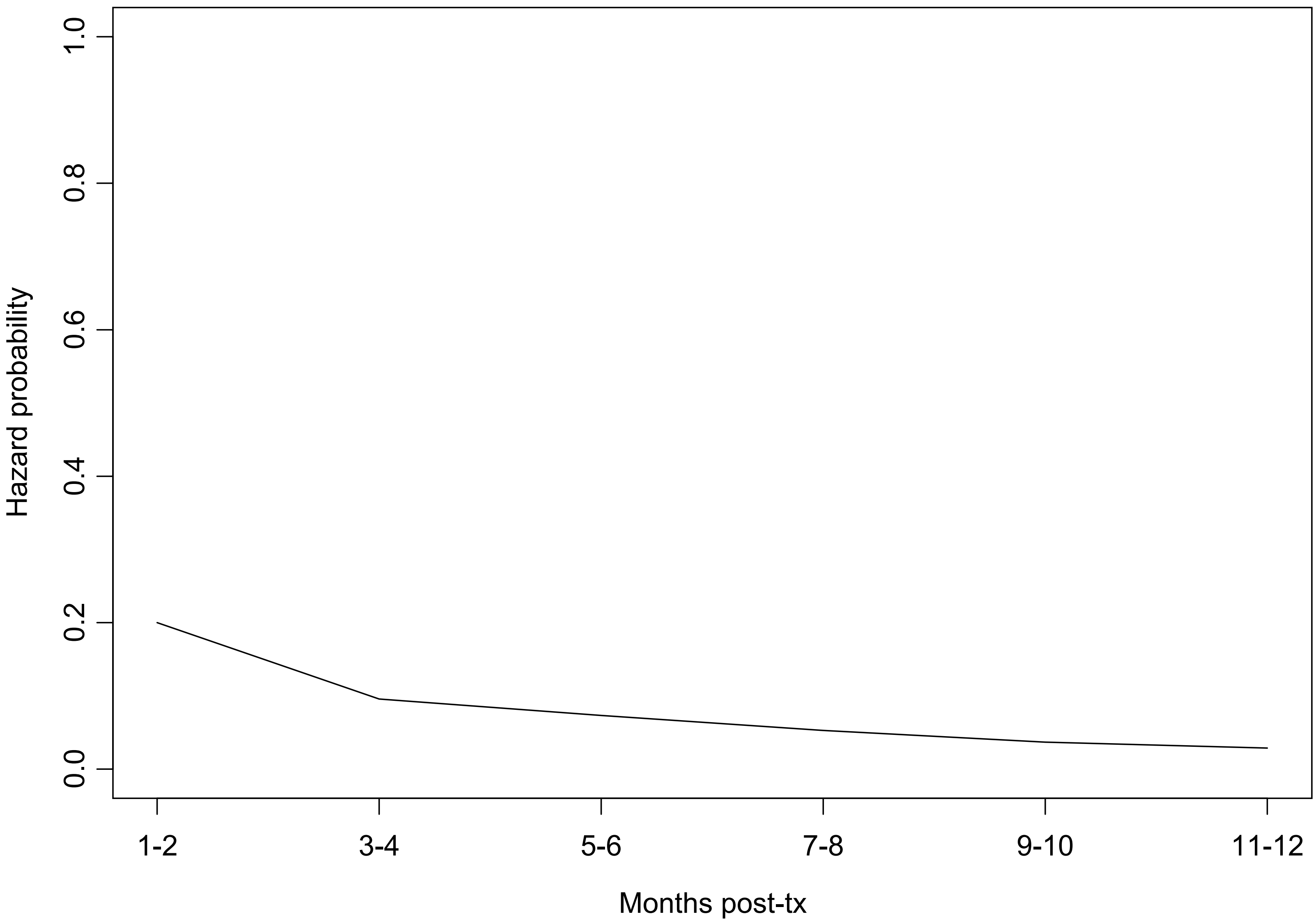
The relationship between $S(j)$ and $h(j)$ is given by

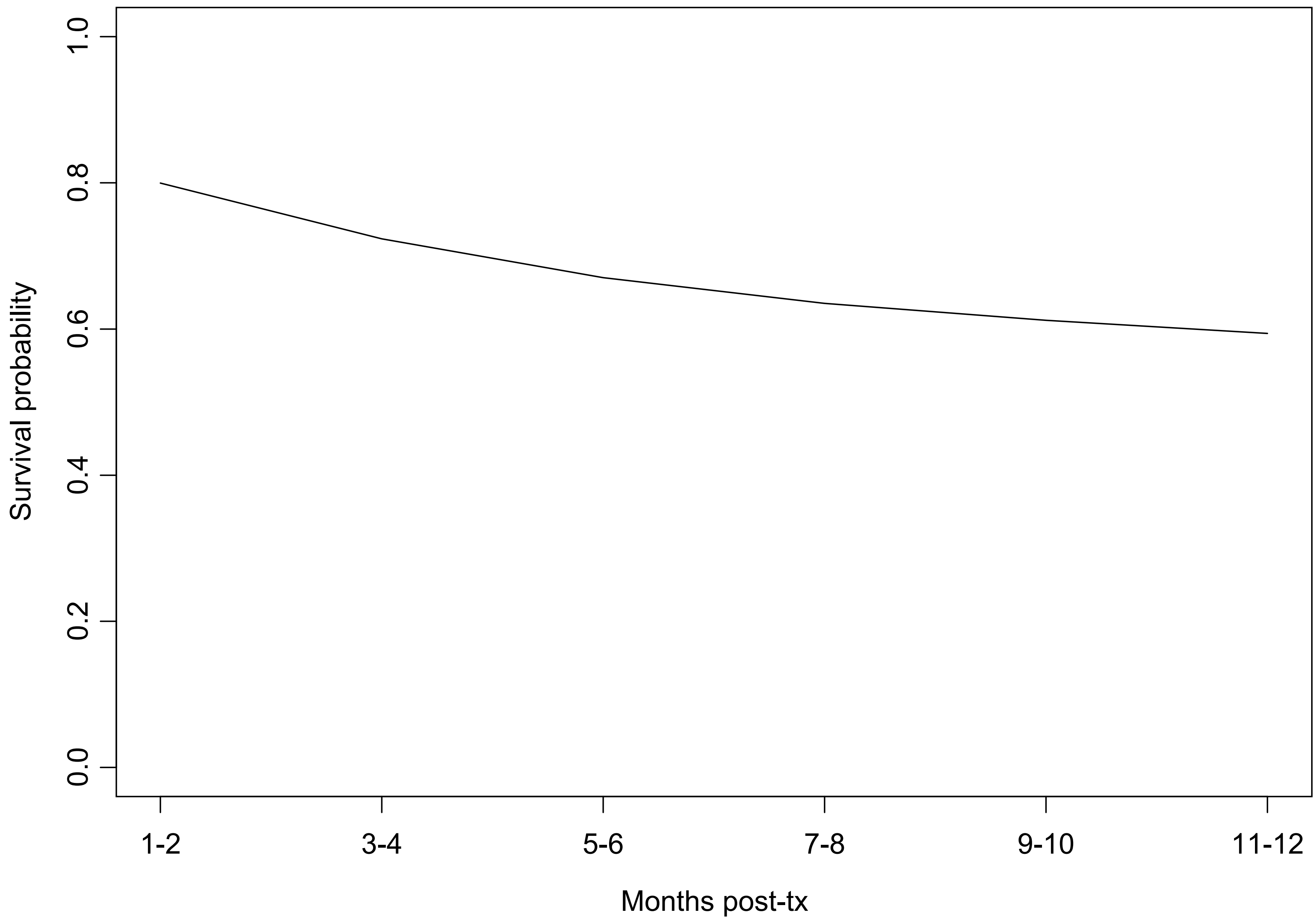
$$S(j) = P(T > j) =$$

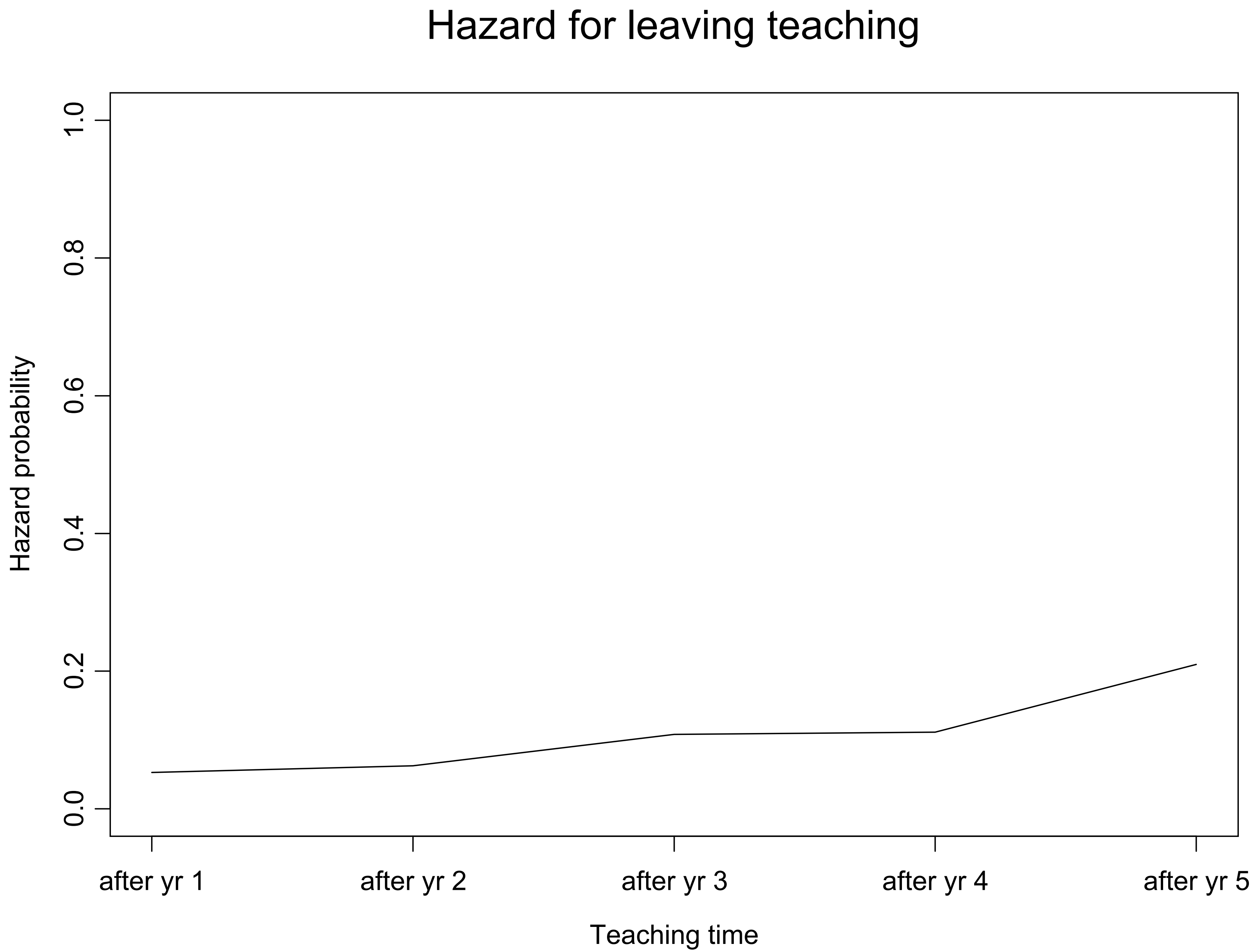
$$P(T > a \mid T \geq a) \times$$
$$P(T > a - 1 \mid T \geq a - 1) \times \ldots$$
$$P(T > 1 \mid T \geq 1) =$$

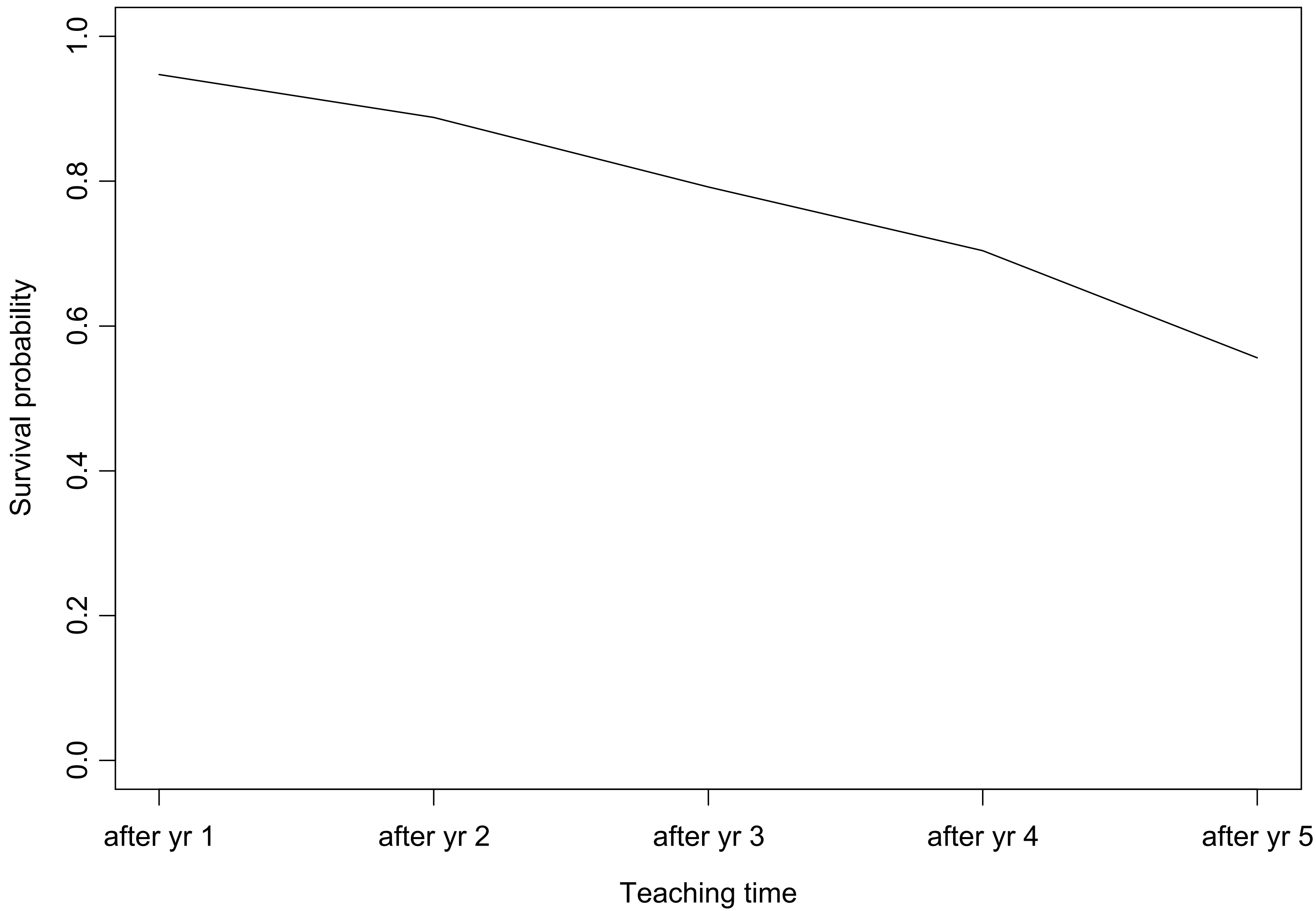
$$\Pi \; [1 - h(k)] \qquad \{k=1 \rightarrow a\}$$

Most survival models are specified in terms of the hazard probabilities.


Hazard probability
0.0
0.2
0.4
0.6
0.8
1.0
1-2
3-4
5-6
7-8
9-10
11-12
Months post-tx


Survival probability
0.0
0.2
0.4
0.6
0.8
1.0
1-2
3-4
5-6
7-8
9-10
11-12
Months post-tx


Hazard for leaving teaching
Hazard probability
0.0
0.2
0.4
0.6
0.8
1.0
after yr 1
after yr 2
after yr 3
after yr 4
after yr 5
Teaching time


Survival probability
0.0
0.2
0.4
0.6
0.8
1.0
after yr 1
after yr 2
after yr 3
after yr 4
after yr 5
Teaching time

Defining risk

- What is the event, i.e., for what is the individual at-risk?
- What defines risk onset, i.e., t=0?
- Under what circumstances does an individual cease to be at-risk?
- Under what circumstances is the event time of an individual unknown or not observed?

Risk for RIA Example

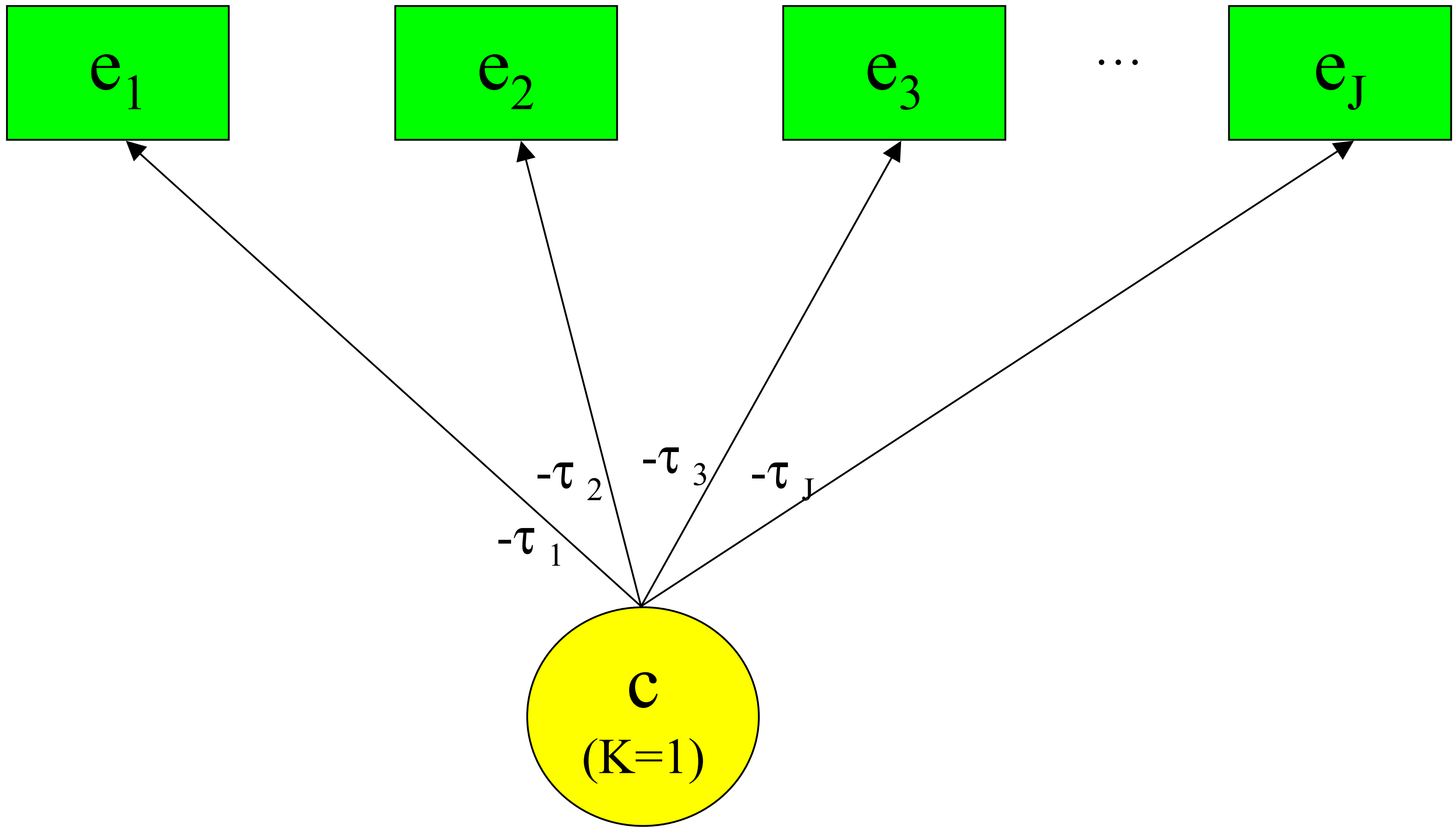
- The event is first post-tx domestic violence episode.
- Risk for all subjects begins at the conclusion of treatment.
- Once the first episode has occurred, an individual is no longer at-risk.
- Event time is unknown if individual is lost to follow-up or the time is greater than 12 months.

Risk for UTEC Example

- The event is first departure from full-time classroom teaching.
- Risk for all subjects begins at the end of the first year of teaching.
- Once an individual has left teaching for the first time, he/she is no longer at-risk.
- Event time is unknown if individual is lost to follow-up or the time is greater than 5 years.

	After Yr 1	*After Yr 2*	*After Yr 3*	*After Yr 4*	*After Yr 5*
At-risk	513	352	240	144	81
Event	27	22	26	16	17
h(j)	0.05	0.06	0.11	0.11	0.21

i	e_1	e_2	e_3	e_4	e_5
1	0	0	1	.	.
2	0	0	.	.	.
3	0	0	0	0	0



$$\hat{h}(j) = \hat{P}(E_j = 1)$$

DTSA Model in Mplus

$$\operatorname{logit} h(j) = \log\left(\frac{h(j)}{1-h(j)}\right) = -\tau_j$$

$$h(j) = \frac{1}{1+\exp(\tau_j)}$$

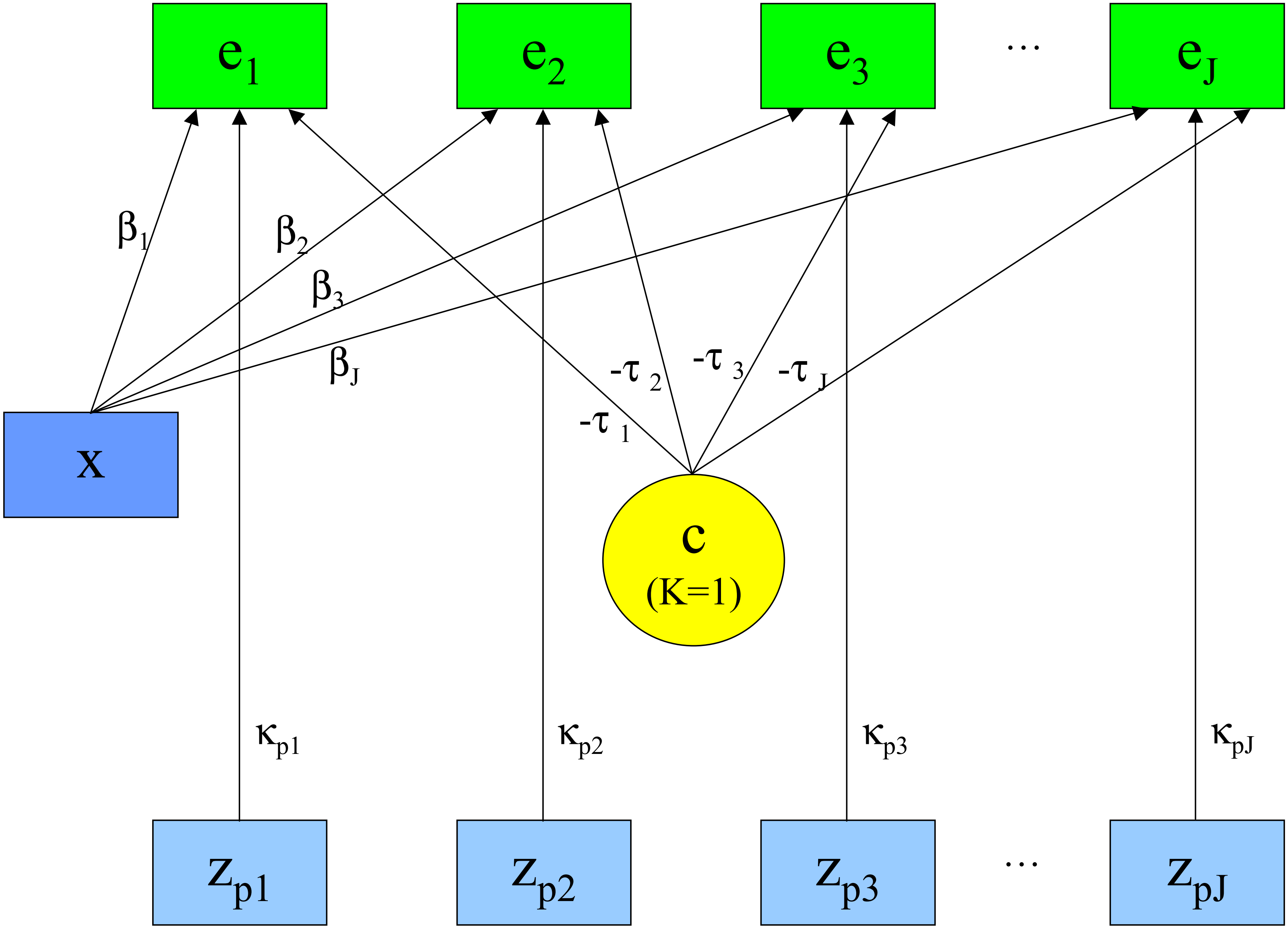
Examples w/o covariates

- Violence_nocov.out

$$\hat{\tau}_1 = 1.386$$

$$\hat{h}(1) = \frac{1}{1+\exp(1.386)} = 0.20$$

- Teacher_nocov.out


e_1
e_2
e_3
...
e_J
β_1
β_2
β_3
β_J
$-\tau_1$
$-\tau_2$
$-\tau_3$
$-\tau_J$
x
c
(K=1)
κ_{p1}
κ_{p2}
κ_{p3}
κ_{pJ}
z_{p1}
z_{p2}
z_{p3}
...
z_{pJ}

DTSA Model w/ covariates*

$$\operatorname{logit} h(j) = \log\left(\frac{h(j)}{1-h(j)}\right) = -\tau_j + \beta_j x + \kappa_j z_j$$

$$h(j) = \frac{1}{1+\exp(\tau_j - \beta_j x - \kappa_j z_j)}$$

* This model, for single events with no random effects, yields identical results to the formulation in the traditional logistic regression model, á la Singer & Willet.

UTEC Example w/ covariates

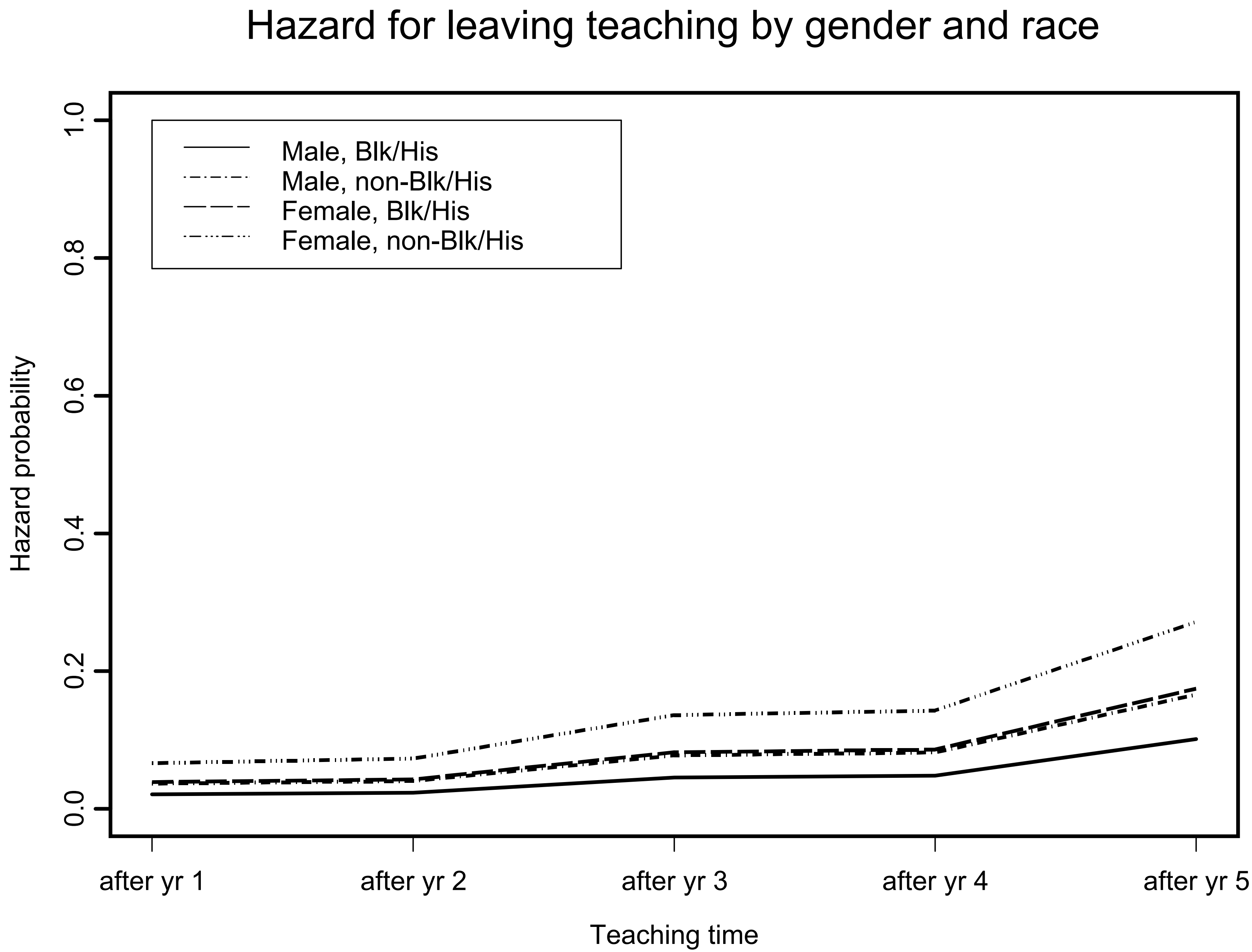
- Teacher_cov.out

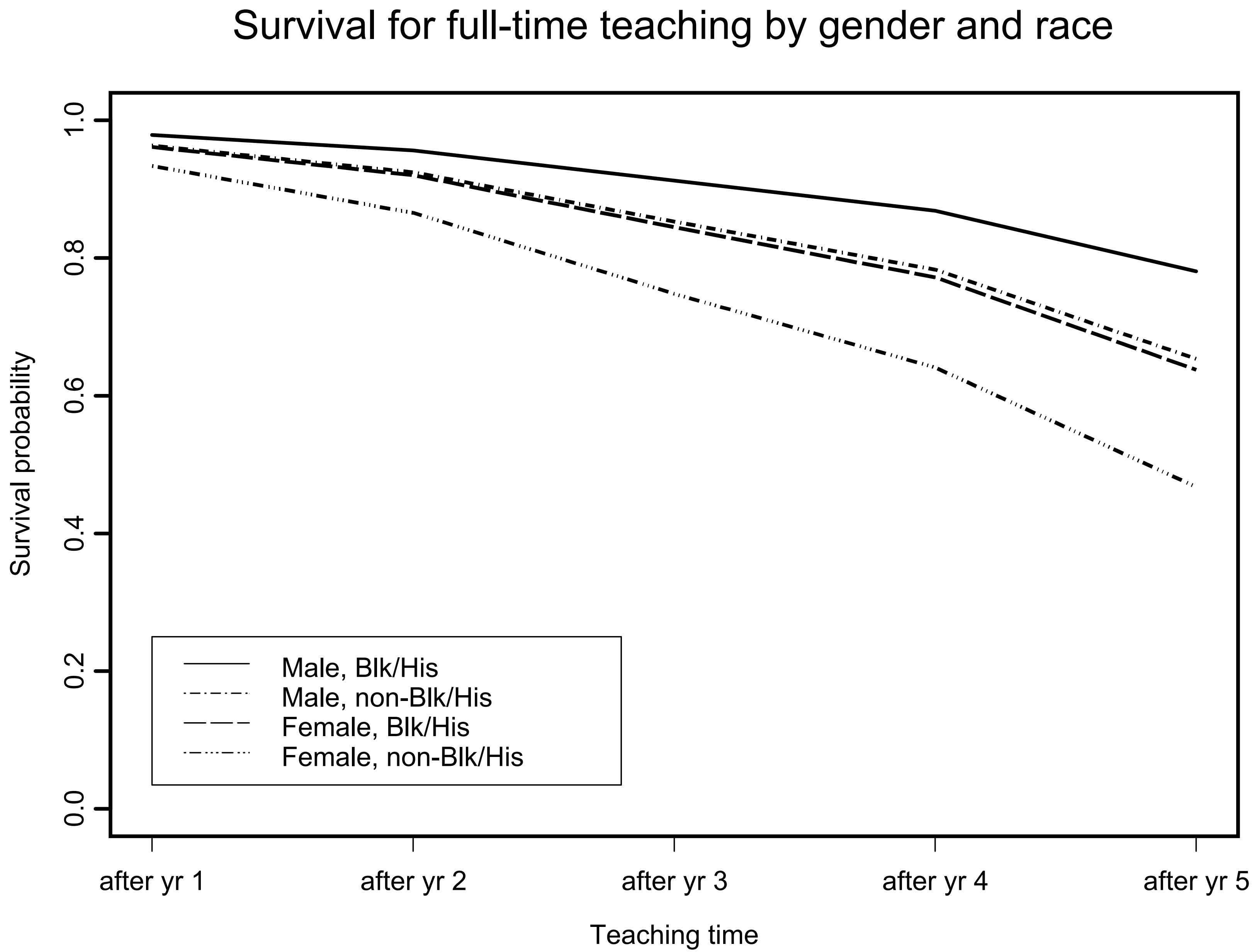
$$\log\left(OR_{gender}\right) = -0.627$$

$$Hazard\ OR = \exp(-0.627) = 0.53$$

$$\hat{h}(j = 2 \mid male, white) =$$

$$\frac{1}{1 + \exp(2.545 - (-0.627)(1) - (-0.568)(0))} = 0.04$$


Hazard for leaving teaching by gender and race
Male, Blk/His
Male, non-Blk/His
Female, Blk/His
Female, non-Blk/His
Hazard probability
0.0
0.2
0.4
0.6
0.8
1.0
after yr 1
after yr 2
after yr 3
after yr 4
after yr 5
Teaching time


Survival for full-time teaching by gender and race
Survival probability
1.0
0.8
0.6
0.4
0.2
0.0
after yr 1
after yr 2
after yr 3
after yr 4
after yr 5
Teaching time
Male, Blk/His
Male, non-Blk/His
Female, Blk/His
Female, non-Blk/His

RIA Example w/ covariates

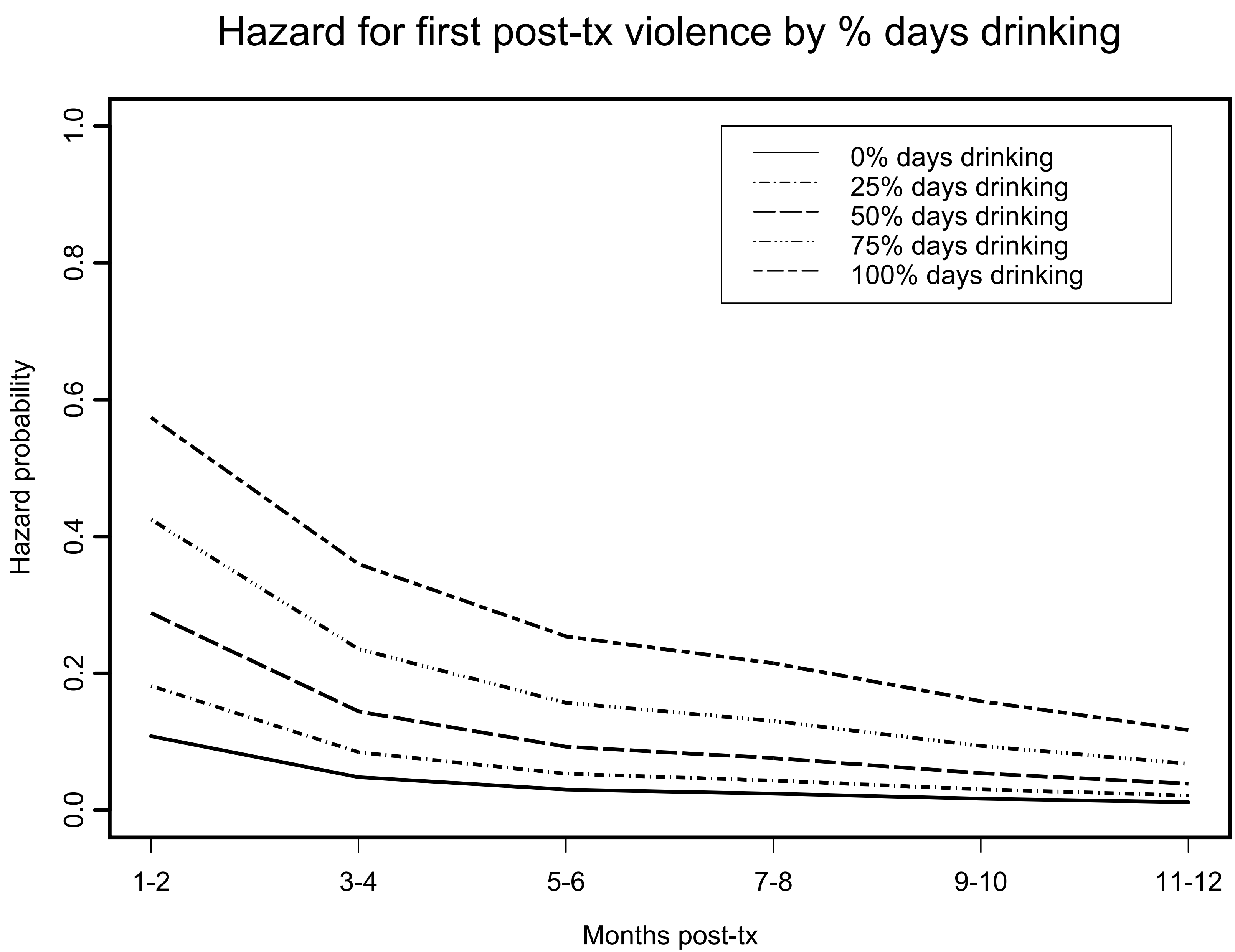
- Violence_cov.out

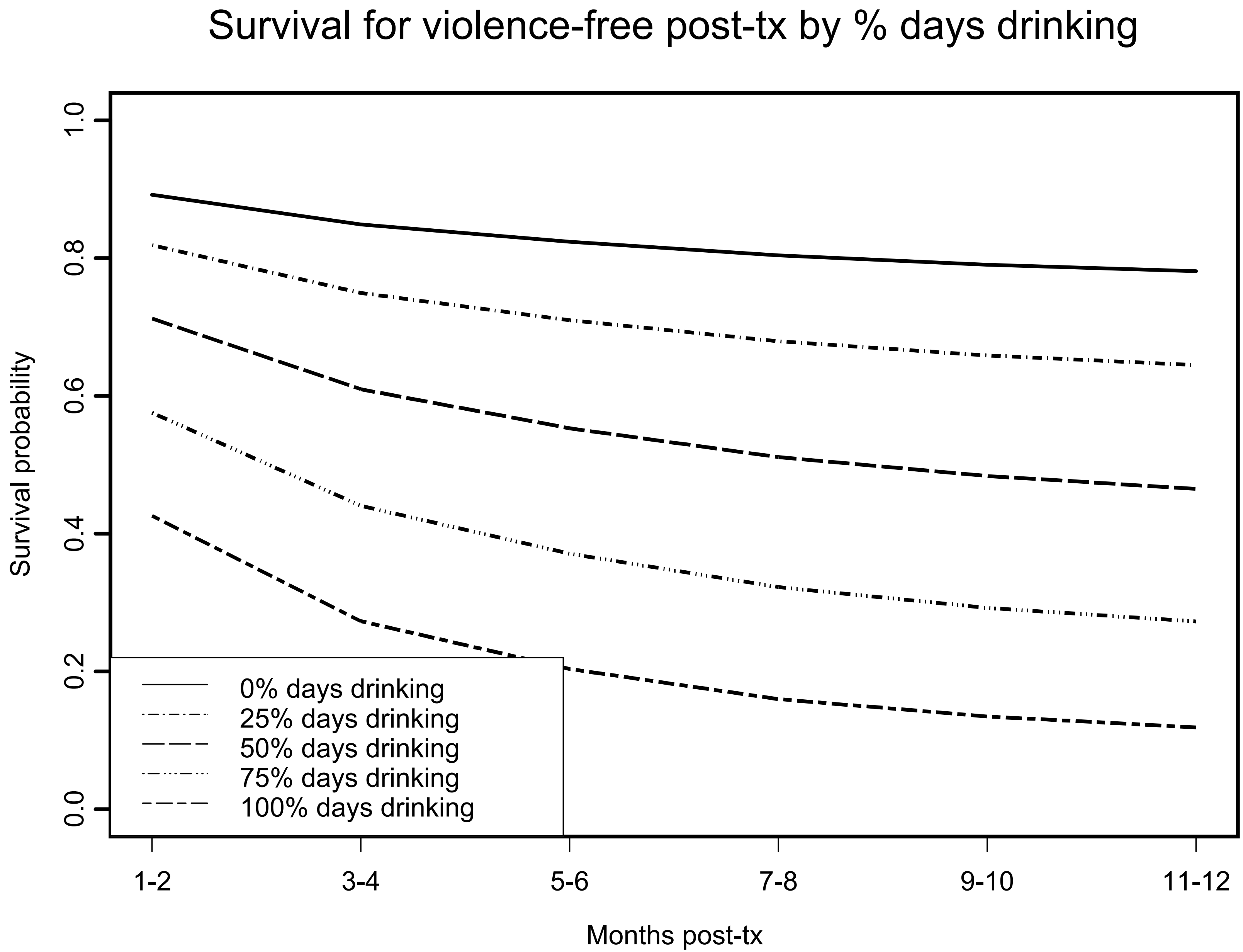
$$\log\left(OR_{pdd}\right) = 2.408$$

$$Hazard\ OR = \exp(2.408) = 11.11$$

$$\hat{h}(j = 1 \mid edw1 = 1, inchi = 0, pdd = 0.25) =$$

$$\frac{1}{1 + \exp(1.589 - (-0.668)(1) - 0 - (2.408)(0.25))} = 0.17$$


Hazard for first post-tx violence by % days drinking
Hazard probability
0.0
0.2
0.4
0.6
0.8
1.0
0% days drinking
25% days drinking
50% days drinking
75% days drinking
100% days drinking
1-2
3-4
5-6
7-8
9-10
11-12
Months post-tx



Survival for violence-free post-tx by % days drinking
Survival probability
0.0
0.2
0.4
0.6
0.8
1.0
0% days drinking
25% days drinking
50% days drinking
75% days drinking
100% days drinking
1-2
3-4
5-6
7-8
9-10
11-12
Months post-tx

Proportional hazard odds*

- Proportional (time-invariant effects):

```
elvtch1-elvtch5 on gender (1);
```

- Non-proportional (time-varying effects):

```
elvtch1-elvtch5 on gender;
```

- Piecewise effects:

```
elvtch1-elvtch2 on gender (1);
elvtch3-elvtch5 on gender (2);
```

*Nested models can be statistically compared using the Likelihood Ratio Test.

Baseline hazard

- Unstructured*:

```
[elvtch1$1-elvtch5$1];
```

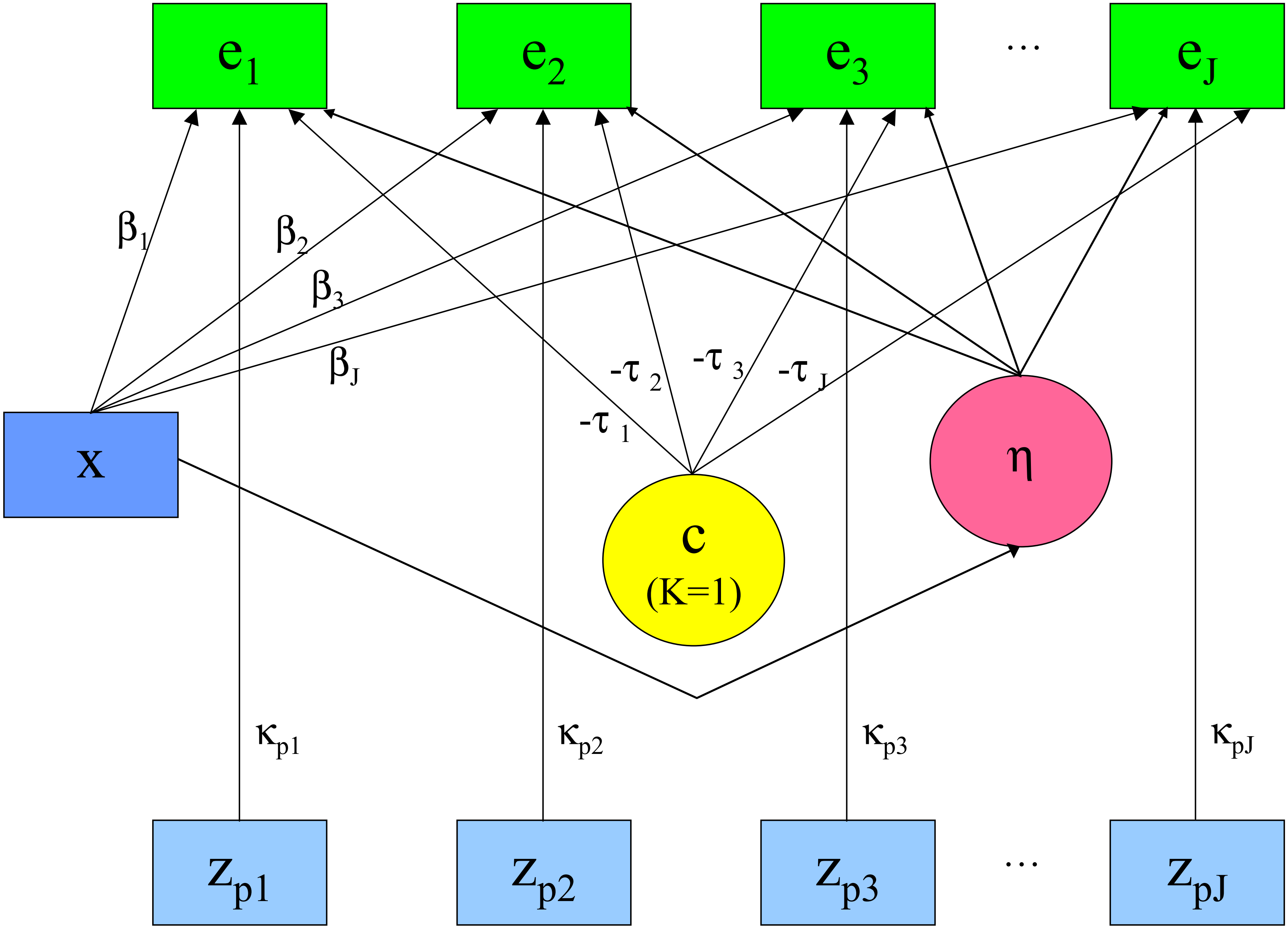
```
*Similar to piecewise in continuous-time
```

- Constant:

```
[elvtch1$1-elvtch5$1] (1);
```

- Piecewise:

```
[elvtch1$1-elvtch2$1] (1);
[elvtch3$1-elvtch5$1] (2);
```



e1
e2
e3
...
eJ
β1
β2
β3
βJ
-τ 1
-τ 2
-τ 3
-τ J
x
c
(K=1)
η
κp1
κp2
κp3
κpJ
zp1
zp2
zp3
...
zpJ

Uses of η (ψ = 0)

Proportional hazard odds:

```
f by elvtch1-elvtch5@1;
f on gender;
```

$$\operatorname{logit} h_i(j) = -\tau_j + \kappa_j z_{ij} + (1)\eta_i$$

$$\eta_i = \omega x_i$$

$$\operatorname{logit} h_i(j) = -\tau_j + \kappa_j z_{ij} + \omega x_i$$

elvtch5	ON	
GENDER		-0.627
BLKHIS		-0.568
Thresholds		
elvtch1$1		2.645
elvtch2$1		2.545
elvtch3$1		1.847
elvtch4$1		1.793
elvtch5$1		0.988

F	ON	
GENDER		-0.627
BLKHIS		-0.568
Intercepts		
F		0.000
Thresholds		
AE32$1		2.645
AE33$1		2.545
AE34$1		1.847
AE35$1		1.793
AE36$1		0.988

Uses of η (ψ = 0)

Baseline hazard time structure:

```
i s | evio11@0 evio12@1
  evio13@2 evio14@3 evio15@4;
i on edw1 inchi;
```

$$\operatorname{logit} h_i(j) = -\tau + \kappa_j z_{ij} + (1)\eta_{0i} + (j-1)\eta_{1i}$$

$$\eta_{0i} = \omega x_i$$

```
I        ON
   EDW1              -0.675    0.277     -2.431
   INCHI             -0.668    0.328     -2.038

Means
   S                 -0.498    0.101     -4.931

Intercepts
   I                  0.000    0.000      0.000

Thresholds
   EVIO11$1           1.718    0.229      7.494
   EVIO12$1           1.718    0.229      7.494
   EVIO13$1           1.718    0.229      7.494
   EVIO14$1           1.718    0.229      7.494
   EVIO15$1           1.718    0.229      7.494
   EVIO16$1           1.718    0.229      7.494
```

Select references

- Allison, P.D. (1984). *Event history analysis: Regression for longitudinal event data*. Newbury Park, CA: Sage Publications, Inc.
- Masyn, K.E. (2003). *Discrete-time survival mixture analysis for single and recurrent events using latent variables.* Unpublished doctoral dissertation, University of California, Los Angeles. http://www.statmodel.com/download/masyndissertation.pdf
- Muthén, B. & Masyn, K. (2003). Discrete-time survival mixture analysis. *Accepted for publication. Currently under revision*.
- Muthén, L., & Muthén, B. (2004). Mplus User's Guide (Version 3.0). Los Angeles: Muthén & Muthén.
- Singer, J.D. & Willett, J.B. (2003). *Applied longitudinal data analysis: Modeling change and event occurrence.* New York: Oxford University Press.
- Vermunt, J.K. (2002). A general latent class approach to unobserved heterogeneity in the analysis of event history data. In J.A. Hagenaars & A.L. McCutcheon (Eds.), *Applied latent class analysis* (pp. 383-407). Cambridge: Cambridge University Press.