Summary of Day 3

An introduction to survival analysis

Mark Stevenson

Faculty of Veterinary and Agricultural Sciences

The University of Melbourne, Parkville Victoria 3010 Australia

mark.stevenson1@unimelb.edu.au

Topics covered

- Baseline hazard
- Model building
- Dealing with non-proportionality of hazards

Baseline hazard

- Easier to model instantaneous risk of an event (i.e. hazard) than it is to model survival

- A parametric model based on the exponential distribution can be parameterised using a linear model for the log-hazard:

$$\log h_i(t) = \alpha + \beta_1 x_{1i} + \beta_2 x_{2i} + \ldots + \beta_m x_{mi}$$

Baseline hazard

- The Cox proportional hazards model removes α (the intercept) and replaces it with $\alpha(t)$:

$$\log h_i(t) = \alpha + \beta_1 x_{1i} + \beta_2 x_{2i} + \ldots + \beta_m x_{mi}$$

$$\log h_i(t) = \alpha(t) + \beta_1 x_{1i} + \beta_2 x_{2i} + \ldots + \beta_m x_{mi}$$

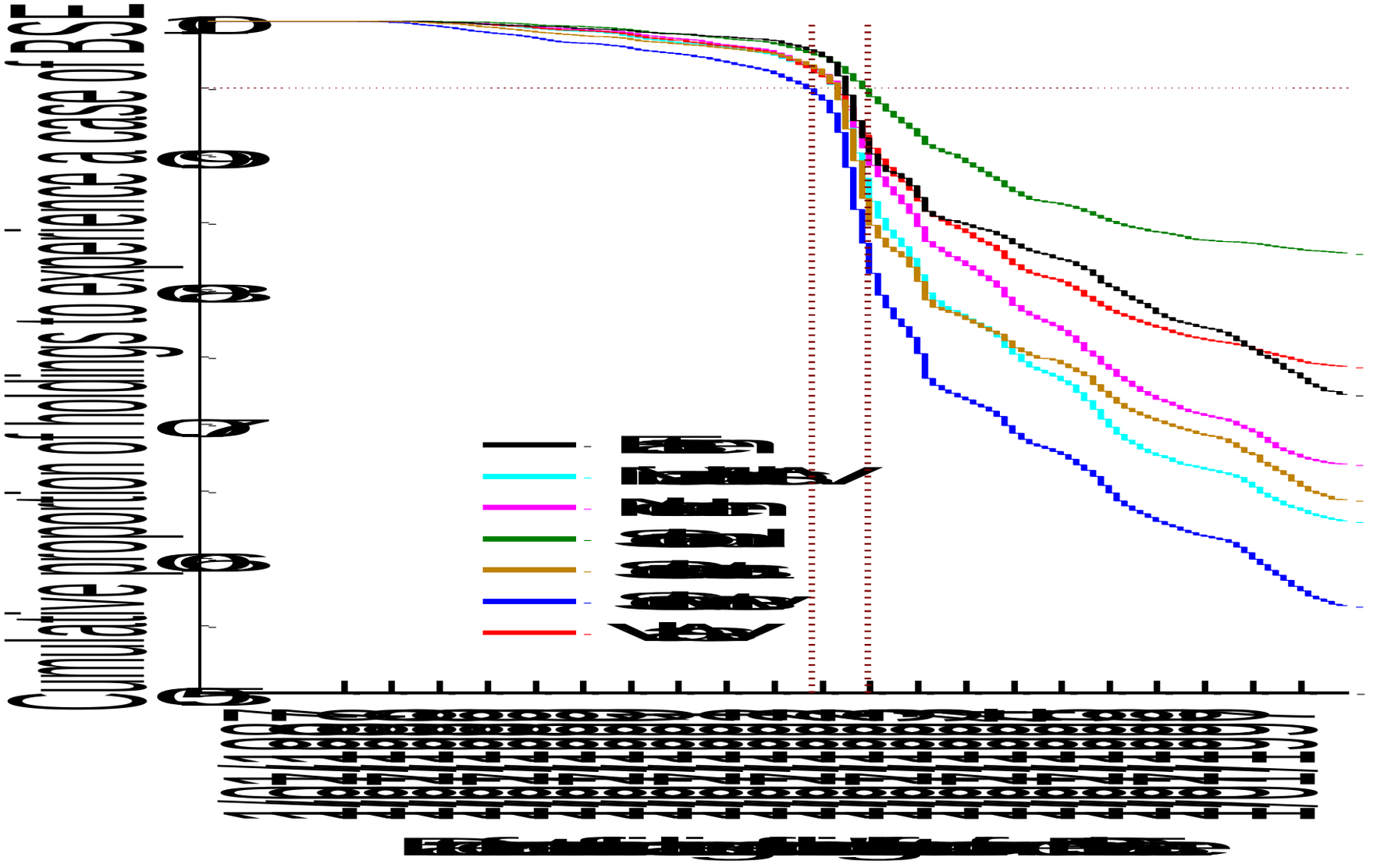
$$h_i(t) = \exp(\alpha(t) + \beta_1 x_{1i} + \beta_2 x_{2i} + \ldots + \beta_m x_{mi})$$

- We can also write:

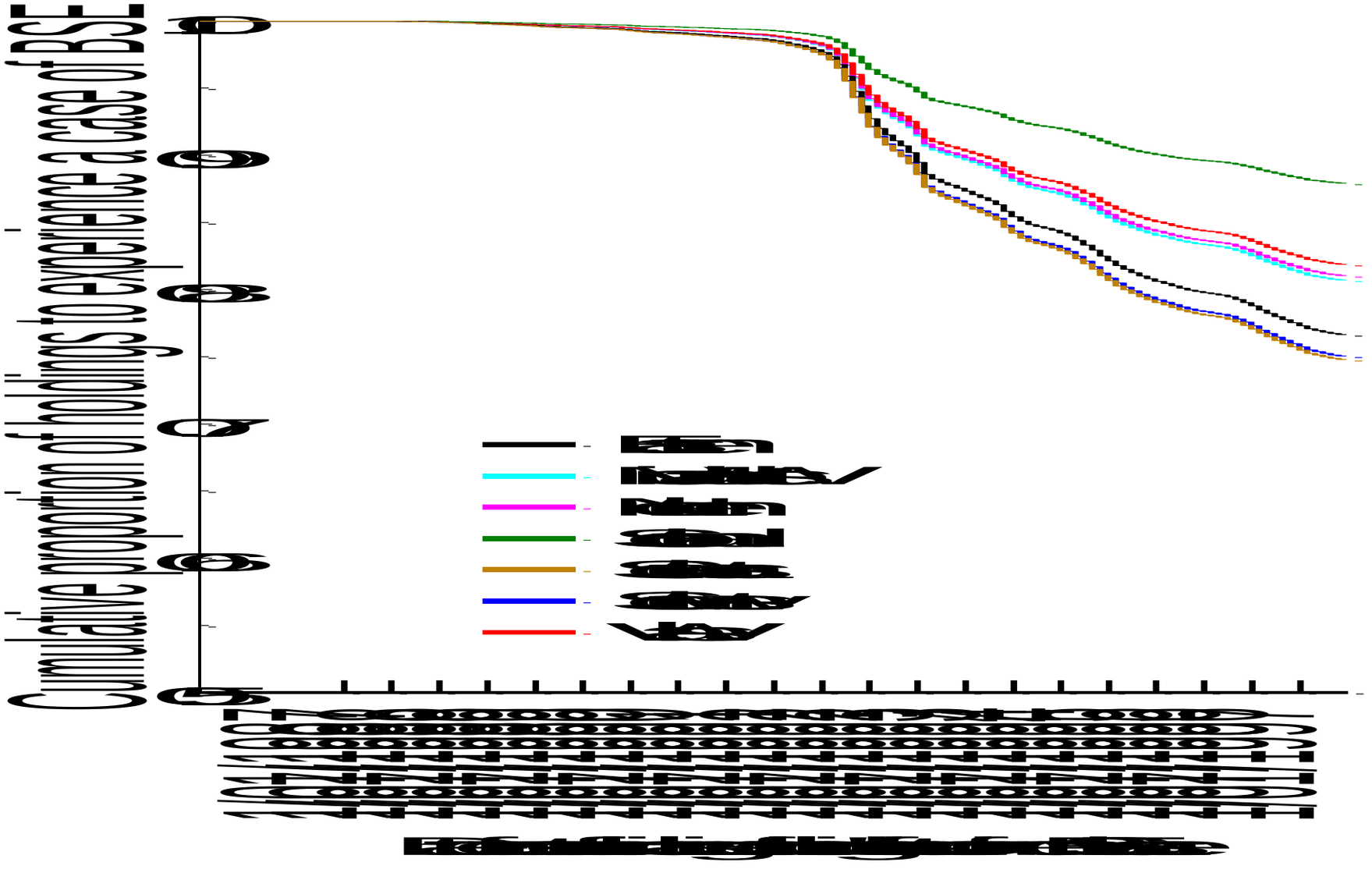
$$h_i(t) = h_0(t)\exp(\beta_1 x_{1i} + \beta_2 x_{2i} + \ldots + \beta_m x_{mi})$$

"Baseline" hazard

Kaplan-Meier survival curves showing time to onset of first case of BSE in British cattle holdings, stratified by region.



Covariate adjusted survival curves showing time to onset of first case of BSE in British cattle holdings, stratified by region.



Topics covered

- Baseline hazard
- Model building
- Dealing with non-proportionality of hazards

Model building

- Essentially the same process as discussed for logistic regression
  - causal diagrams
  - unconditional associations
  - assessment of correlation among candidate explanatory variables
  - model building
  - assessment of confounding
  - interaction

Model building

- Essentially the same process as discussed for logistic regression
  - causal diagrams
  - unconditional associations
  - assessment of correlation among candidate explanatory variables
  - model building
  - check the proportional hazards assumption
  - assessment of confounding
  - interaction

Topics covered

- Baseline hazard
- Model building
- Dealing with non-proportionality of hazards

Dealing with non-proportionality of hazards

- Options
  1. Stratification
  2. Introduce a time-dependent covariate

Dealing with non-proportionality of hazards

- Stratification
  - introduce a separate baseline hazard function for each level of strata that violates the proportional hazard assumption
  - fixes the problem, but by doing this we can't obtain a hazard ratio for the stratified variable since its effect gets 'absorbed' into the baseline hazard function

```
addict.cph04 <- coxph(Surv(stop, status, type = "right") ~
strata(clinic) + prison + dose, method = "breslow", data = dat);
summary(addict.cph04)
```

```
Call:
coxph(formula = Surv(stop, status, type = "right") ~ strata(clinic)
+ prison + dose, data = dat, method = "breslow")

          coef exp(coef) se(coef)      z        p
prison1  0.376     1.457  0.16889  2.23 2.6e-02
dose    -0.035     0.966  0.00645 -5.42 5.9e-08

        exp(coef) exp(-coef) lower .95 upper .95
prison1     1.457      0.686     1.046     2.029
dose        0.966      1.036     0.953     0.978

Rsquare= 0.131   (max possible= 0.994 )
Likelihood ratio test= 33.5  on 2 df,   p=5.29e-08
Wald test            = 32.3  on 2 df,   p=9.73e-08
Score (logrank) test = 33.0  on 2 df,   p=6.98e-08
```

Dealing with non-proportionality of hazards

- Time dependent covariates
  - if the covariate is fixed (i.e. it does not vary with time, but its effect varies with time) we can explore this time-dependent effect by dividing the follow up period into distinct intervals
  - we then fit proportional hazards models to the survival in each interval and compare the coefficients for each covariate across the different time intervals
  - if the coefficient change with time, we have evidence of non-proportional hazards
  - thus, the diagnostic for non-proportionality of hazards is also the solution

```
library(survival); setwd("D:\\TEMP");
dat <- read.table("addict.csv", header = TRUE, sep = ",");
head(dat);
```

```
id start stop status clinic prison dose
1     0  428      1      1      0   50
2     0  275      1      1      1   55
3     0  262      1      1      0   55
4     0  183      1      1      0   30
5     0  259      1      1      1   65
6     0  714      1      1      0   55
```

Reformat the data:

```
id start stop status clinic prison dose
1     0  365      0      1      0   50
1   365  428      1      1      0   50
2     0  275      1      1      1   55
3     0  262      1      1      0   55
4     0  183      1      1      0   30
5     0  259      1      1      1   65
```

Recode the `clinic` variable to make Clinic 2 (the better performing clinic) the reference category:

```
dat$clinic <- as.vector(ifelse(dat$clinic == 2, 0, 1));
```

First of all we'll consider the period before 365 days. Create a new variable called `t1` such that `t1=1` if the time to event is less than or equal to 365 days and zero otherwise:

```
t1 <- rep(0, length(dat[,1]));
t1[dat$stop <= 365] <- 1;
dat <- cbind(dat, t1);
```

Using this coding, the reported hazard for the `clinic * t1` interaction will be for Clinic 1 when time is less than or equal to 365 days.

Next consider the period after 365 days. Create a new variable called `t2` such that `t2 = 1` if the time to event is greater than 365 days and one otherwise:

```
t2 <- rep(0, length(dat[,1]));
t2[dat$stop > 365] <- 1;
dat <- cbind(dat, t2);
```

Using this coding, the reported hazard for the `clinic * t2` interaction will be for Clinic 1 when time is greater than 365 days.

```
head(dat)

id start stop status clinic prison dose t1 t2
1      0  365      0      1      0   50  1  0
1    365  428      1      1      0   50  0  1
2      0  275      1      1      1   55  1  0
3      0  262      1      1      0   55  1  0
4      0  183      1      1      0   30  1  0
5      0  259      1      1      1   65  1  0
```

Now fit the model:

```
addict.cph05 <- coxph(Surv(start, stop, event = status, type = '
counting') ~ prison + dose + I(clinic * t1) + I(clinic * t2), method
= 'breslow', data = dat);
summary(addict.cph05);
```

Variable	Subjects	Failed	Coefficient (SE)	P	Hazard ratio (95%)
Prison:				0.03	
Absent	127	81	-		1.00
Present	111	69	0.3650 (0.1684)		1.44 (1.04 - 2.00)
Dose	238	150	-0.0353 (0.0064)	< 0.01	0.96 (0.95 - 0.98)
Clinic × t1	118	87	0.4802 (0.2548)	0.06	1.62 (0.98 - 2.66) [a]
Clinic × t2	120	63	1.8103 (0.3861)	< 0.01	6.11 (2.87 - 13.03)

[a] Interpretation: compared with the reference category (patients from Clinic 2) when days on treatment is less than 365, after adjusting for the effect of methadone dose and prison record, patients from Clinic 1 had 1.62 (95% CI 0.98 - 2.66) times the daily hazard withdrawing from the treatment program.

Dealing with non-proportionality of hazards

- Conclusions
  - when days on treatment is less than 365 days, patients from Clinic 1 have a 1.62 (95% CI 0.98 to 2.66) times increased hazard of relapse compared with patients from Clinic 2
  - when days on treatment is greater than 365 days, patients from Clinic 1 have a 6.11 (95% CI 2.87 to 13.0) times increased hazard of relapse compared with patients from Clinic 2
  - once you've been at Clinic 1 for greater than 12 months your hazard of relapse increases markedly

THE UNIVERSITY OF

MELBOURNE