Summary of Day 2

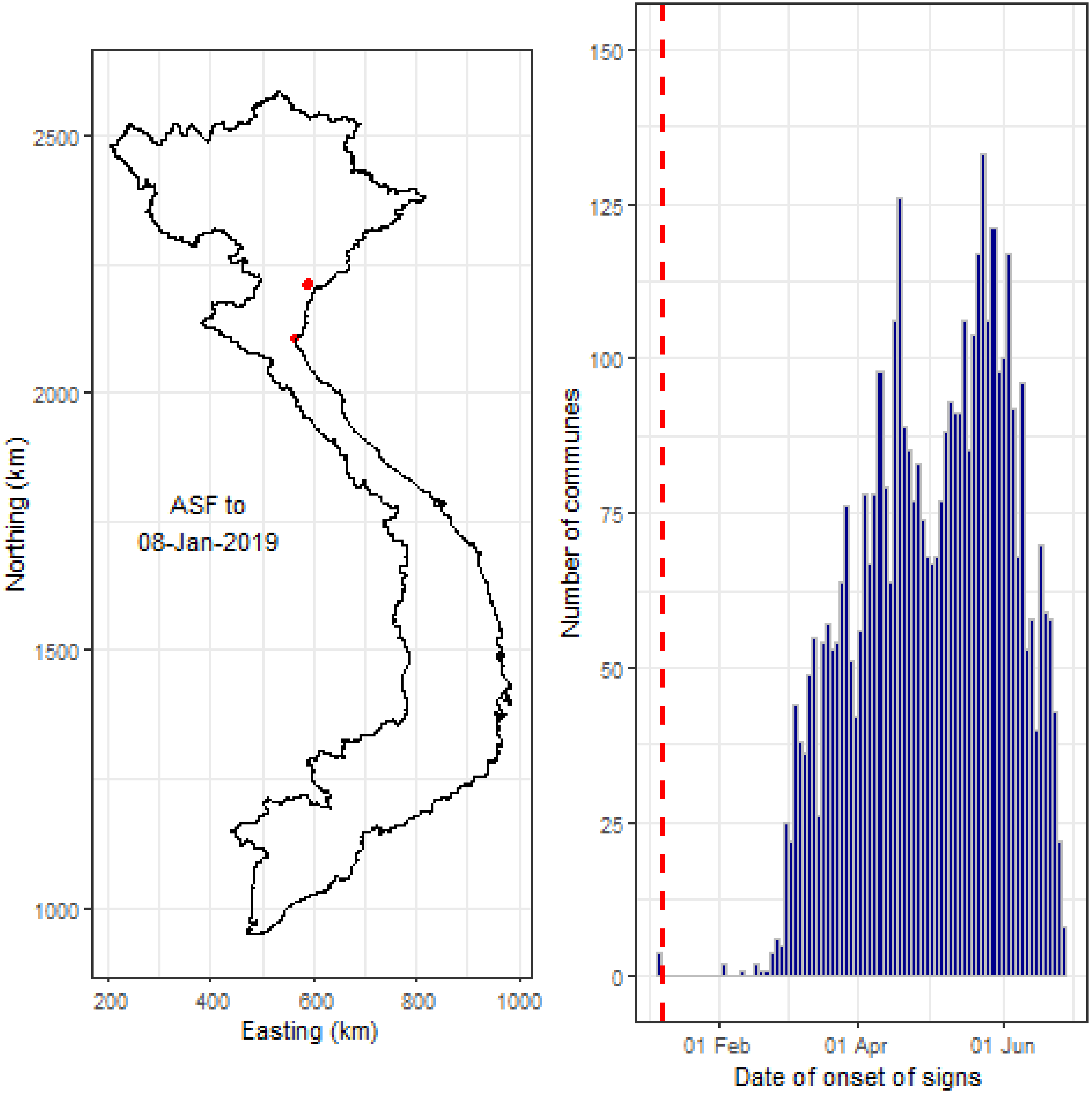
An introduction to survival analysis

Mark Stevenson

Faculty of Veterinary and Agricultural Sciences

The University of Melbourne, Parkville Victoria 3010 Australia

mark.stevenson1@unimelb.edu.au



ASF to
08-Jan-2019
Northing (km)
2500
2000
1500
1000
200
400
600
800
1000
Easting (km)
Number of communes
150
125
100
75
50
25
0
01 Feb
01 Apr
01 Jun
Date of onset of signs

Topics covered

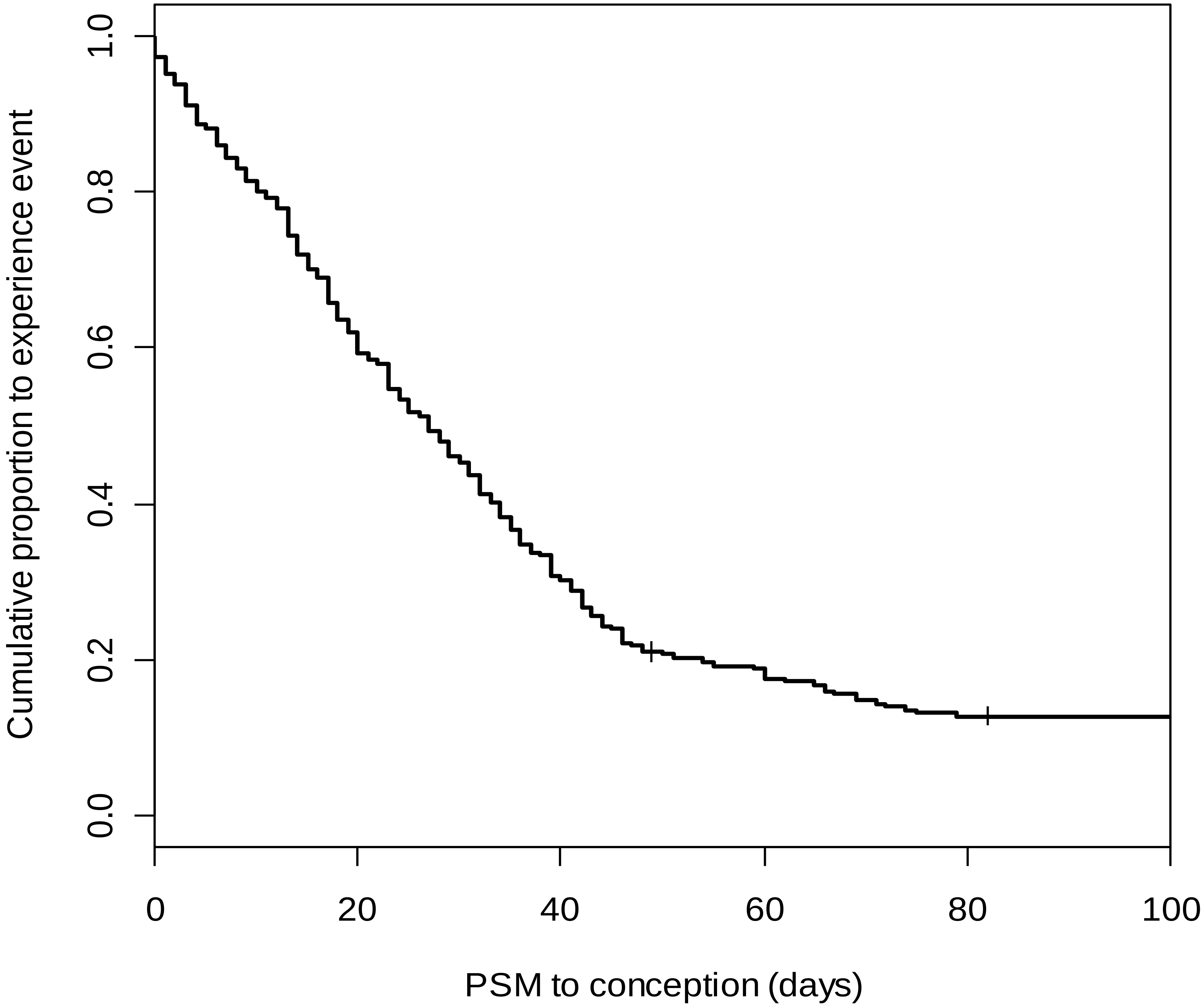
- Parametric and non-parametric survival
- Comparing time to event among groups

Parametric survival distributions

- In the last lecture we talked about ways to describe survivorship
  - x axis: time
  - y axis: proportion of group that had not experienced event

- Some survival functions ‘predictable’ …

Days to conception as a function of days after Planned Start of Mating date in dairy cows.

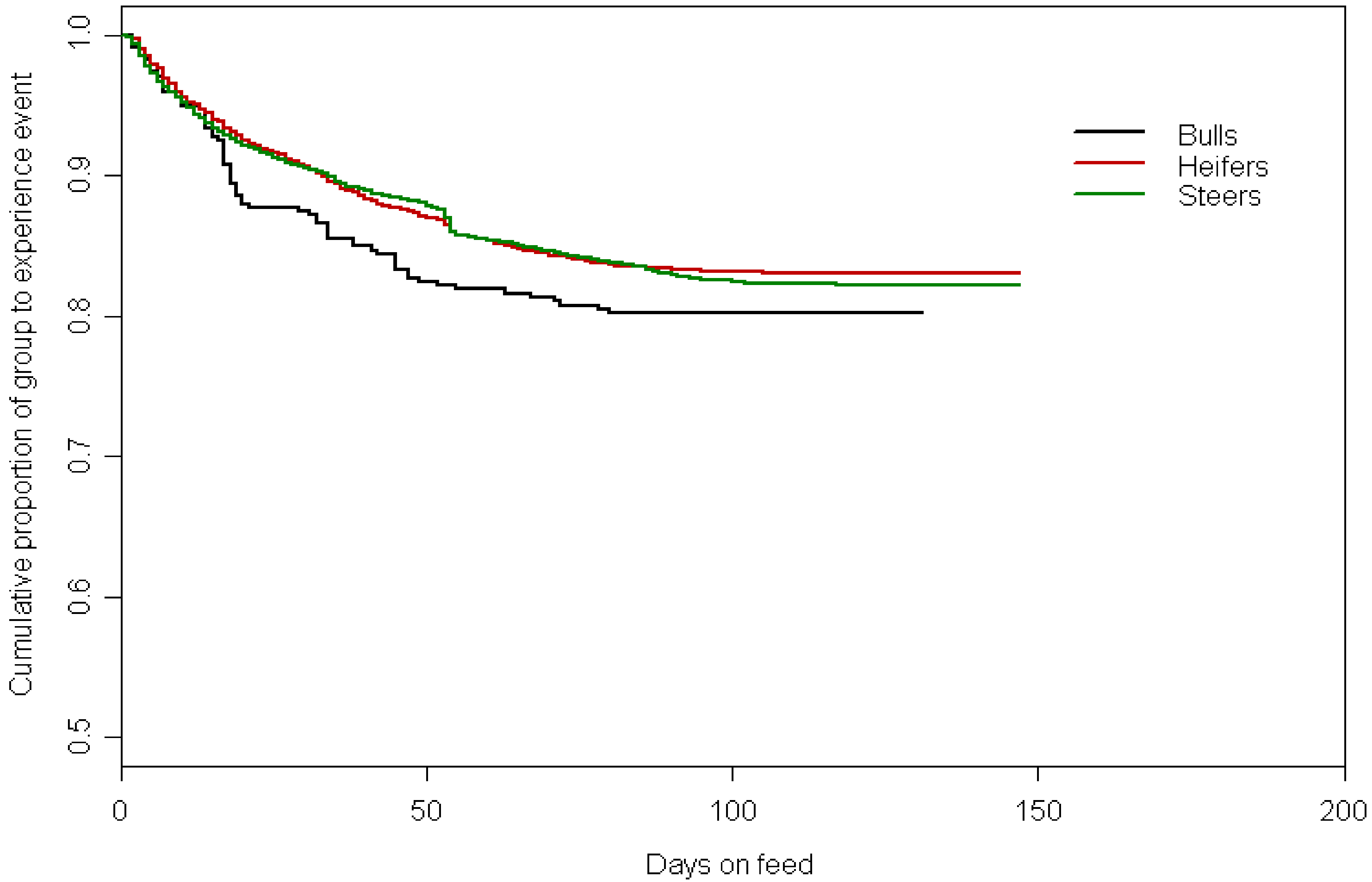


Parametric survival distributions

- Some survival functions ‘unpredictable’ …

Indonesian feedlot.

Kaplan-Meier survival curve showing cumulative proportion of animals pulled for all reasons as a function of calendar date, by stock class.



Parametric survival distributions

- Those distributions follow a nice, well-defined pattern can be described by <u>parametric</u> distributions

- Those distributions that are irregular in shape can only be described by <u>non-parametric</u> distributions

Parametric survival functions

- Several parametric distributions are available, each characterised by different hazard functions:

Distribution	h(t)	f(t)	S(t)
Exponential	λ	$\lambda \exp[-\lambda t]$	$\exp[-\lambda t]$
Weibull	$\lambda p t^{p-1}$	$\lambda p t^{p-1} \exp[-\lambda t^p]$	$\exp[-\lambda t^p]$
Log-logistic	$[ab(at)^{b-1}] / [1 + (at)^b]$	$[ab(at)^{b-1}] / [1 + (at)^b]^2$	$[1 + (at)^b]^{-1}$

$$f(t) = h(t) \times S(t)$$

THE UNIVERSITY OF MELBOURNE

Non-parametric survival functions

- Kaplan-Meier method
  - also known as the Product Limit method
  - based on individual survival times and assumes that censoring is independent of survival time (that is, the reason an observation is censored is unrelated to the cause of failure)

Non-parametric survival functions

- Kaplan-Meier method:

Time	Start n_i	Fail d_i	Censor w_i	At risk $r_i = (n_i – w_i)$	Surv prob $p_i = 1 – [d_i \div (n_i – w_i)]$	Cum surv
0	31	2	3	31-3 = 28	0.93	0.93 × 1.00 = 0.93
1	26	1	2	26-2 = 24	0.96	0.96 × 0.93 = 0.89
2	23	1	2	23-2 = 21	0.95	0.95 × 0.89 = 0.85
3	20	1	2	20-2 = 18	0.94	0.94 × 0.85 = 0.80
etc						

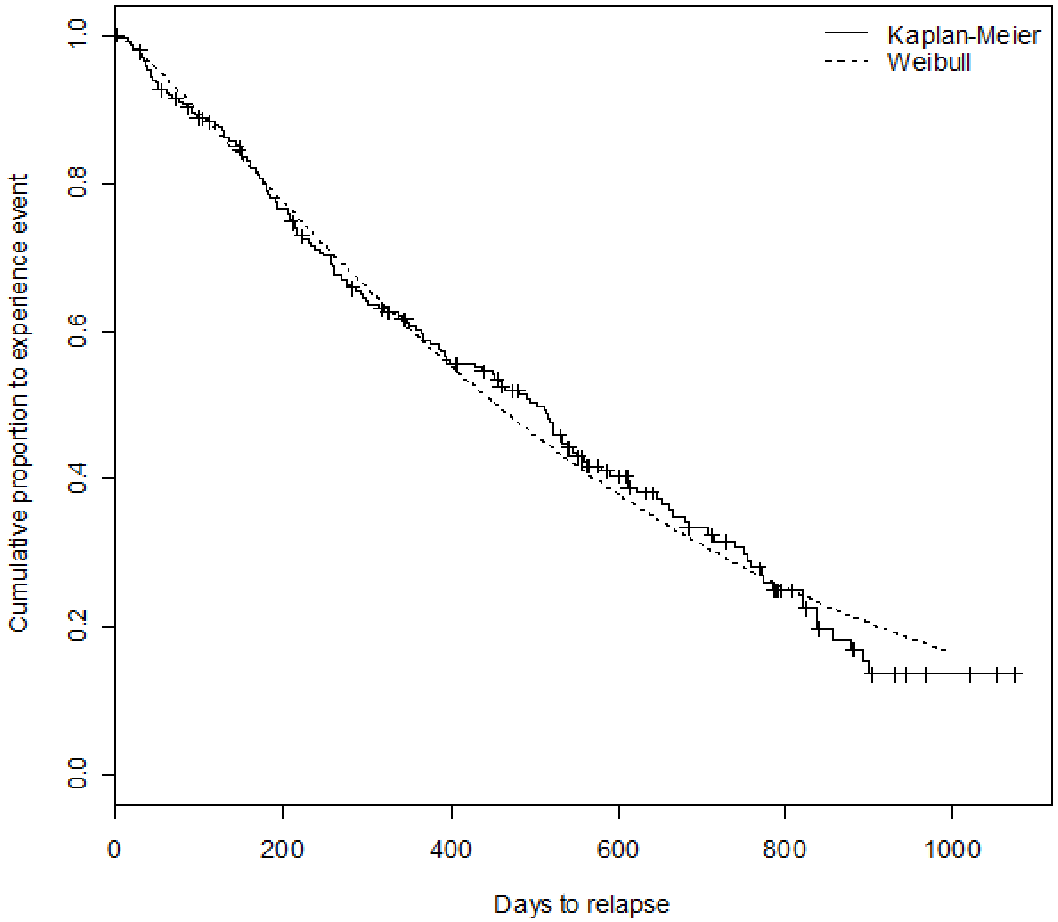
Comparison of Kaplan-Meier and Weibull estimates of survival:

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

addict.we <- survreg(Surv(stop, status) ~ 1, dist = "weib", data =
dat)
addict.km <- survfit(Surv(stop, status) ~ 1, conf.type = "none",
type = "kaplan-meier", data = dat)
```

Using the Weibull distribution μ (the intercept) = -log(λ) and σ (scale) = 1 / p. Thus the scale parameter λ = exp(-μ) and p = 1 / σ. See Tableman and Kim p 78.

```
p <- 1 / addict.we$scale
lambda <- exp(-addict.we$coeff[1])
t <- 1:1000
St <- exp(-(lambda * t)^p)
addict.we <- as.data.frame(cbind(t = t, St = St))
```




```
plot(addict.km, xlab = "Days to relapse", ylab = "Cumulative
proportion to experience event")
lines(addict.we$t, addict.we$St, lty = 2)
legend(x = "topright", legend = c("Kaplan-Meier", "Weibull"), lty =
c(1,2), bty = "n")
```

Parametric and non-parametric survival

- Always check to see if your data are consistent with a parametric distribution

- If yes, give serious thought to the use of parametric methods for analysis: greater statistical power

THE UNIVERSITY OF MELBOURNE

Topics covered

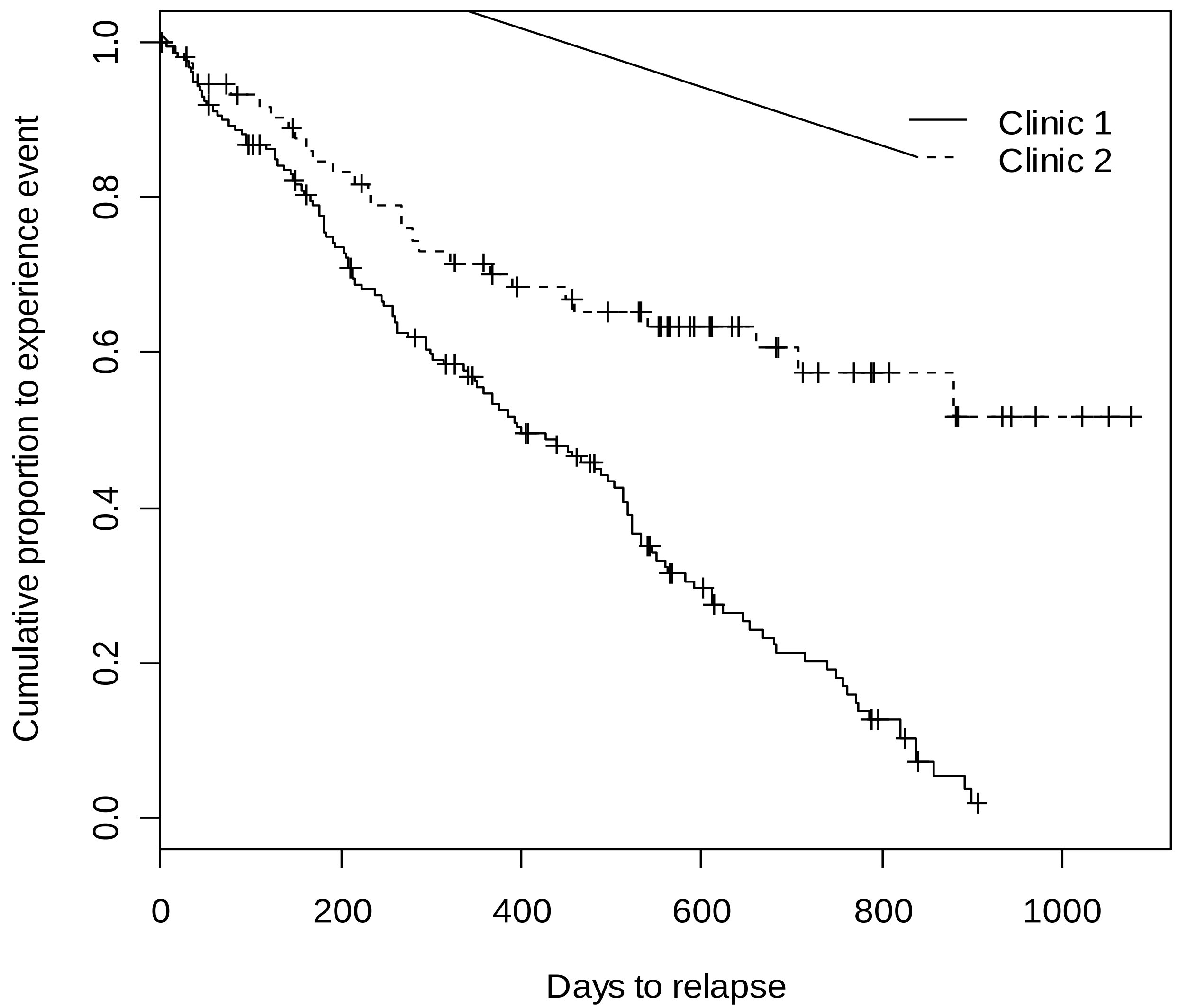
- Parametric and non-parametric survival
- Comparing time to event among groups

Topics covered

- Parametric and non-parametric survival
- Comparing time to event among groups

The log-rank test

- Most commonly-used test for comparing survival distributions
- Applicable to data where there is progressive censoring
- Gives equal weight to early and late failures
- Assumes that hazard ratios between the two groups are parallel




```
library(survival); setwd("D:\\TEMP");
dat <- read.table("addict.csv", header = TRUE, sep = ",")
addict.km <- survfit(Surv(stop, status) ~ clinic, type = "kaplan-
meier", data = dat)
plot(addict.km, xlab = "Days to relapse", ylab = "Cumulative
proportion to experience event", lty = c(1,2), legend.text =
c("Clinic 1", "Clinic 2"), legend.pos = 1, legend.bty = "n")
```

```
survdiff(Surv(stop, status) ~ clinic, data = dat, na.action =
na.omit, rho = 0)

Call:
survdiff(formula = Surv(stop, status) ~ clinic, data = dat, na.action = na.omit,
rho = 0)

           N Observed Expected (O-E)^2/E (O-E)^2/V
clinic=1 163      122     90.8      10.7      28.1
clinic=2  75       28     59.2      16.4      28.1

 Chisq = 28.1  on 1 degrees of freedom, p = 1.18e-07
```