Survival analysis:
Comparing time to event among groups

Mark Stevenson

Faculty of Veterinary and Agricultural Sciences

The University of Melbourne, Parkville Victoria 3010 Australia

mark.stevenson1@unimelb.edu.au

Roadmap

- Background
- Log-rank test
- Breslow's test
- Others

Background

- Often we want to compare survivorship of one group with another
  - did animals survive longer in one herd compared with another?
  - did disease take longer to develop in one region of a country compared with another?
  - did patients survive longer after one therapy compared with another?

Background

- Survival curves for different levels of an explanatory variable provide an effective screening method to identify factors influential in determining survival

- Once influential factors are screened using these methods their effect can be more comprehensively tested using a multivariate analysis

Background

- When there are no censored observations, standard non-parametric tests can be used to compare two survival distributions
  - if the groups are independent, a Wilcoxon or Mann-Whitney U test can be used
  - if the groups are dependent a Sign Test can be used

Background

- When there are censored observations, the following tests can be used:
  - log-rank test (≡ Mantel log-rank test, Cox Mantel log-rank test, Mantel-Haenszel test)
  - Breslow's test (≡ Gehan's generalised Wilcoxon test)
  - Cox Mantel test
  - Peto and Peto modification of the Gehan-Wilcoxon test
  - Cox's F test
  - Tarone Ware test

Roadmap

- Background
- Log-rank test
- Breslow's test
- Others

The log-rank test

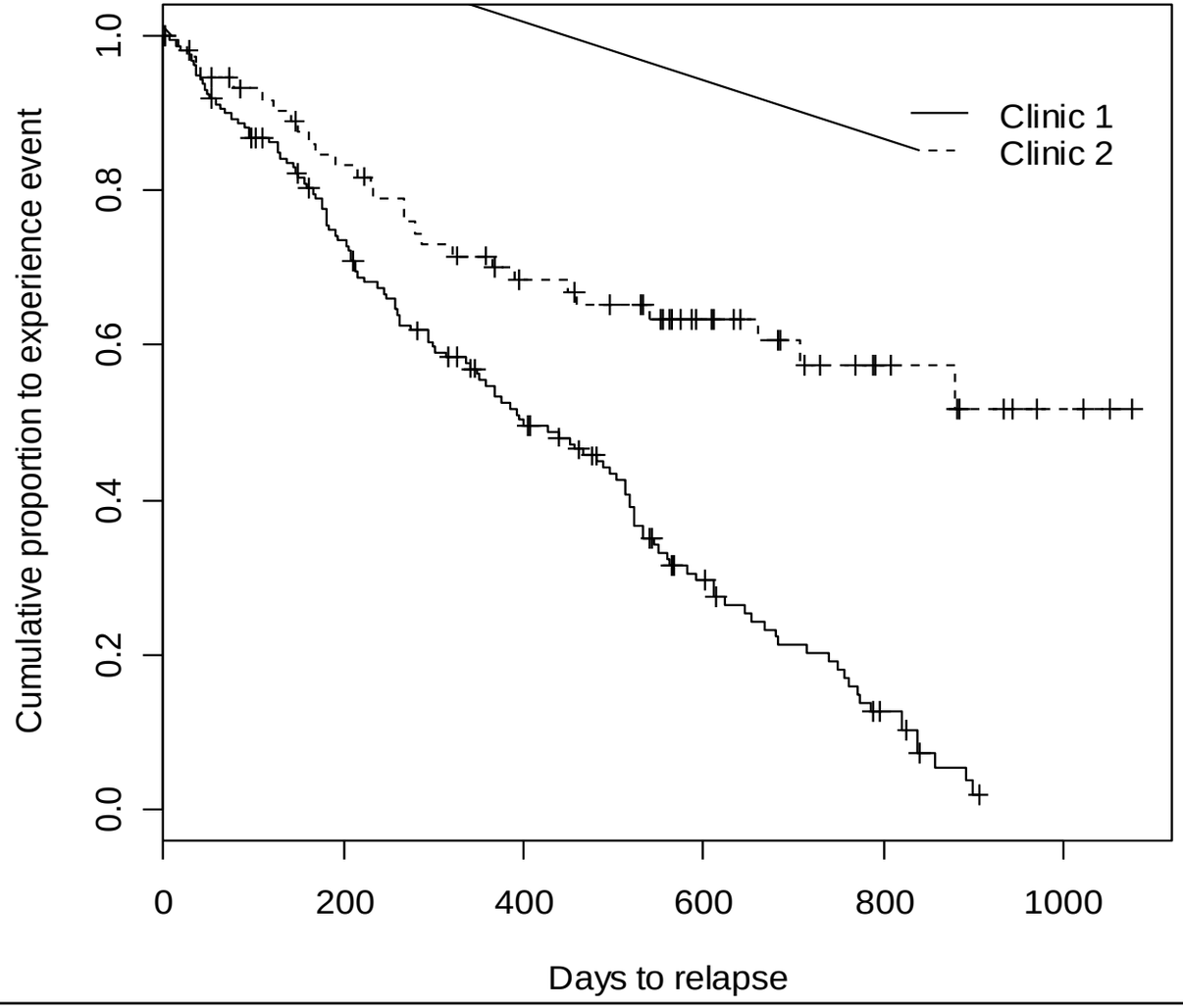
- Most commonly-used test for comparing survival distributions
- Applicable to data where there is progressive censoring
- Gives equal weight to early and late failures
- Assumes that hazard ratios between the two groups are parallel

The log-rank test

- Method:
  - takes each time point when an event occurs and generates a 2×2 table showing the number of events and the total number of subjects under follow up
  - for each table the observed events in one group, the expected events and the variance of the expected number of events are calculated
  - these quantities are summed over all tables to yield a χ^2 statistic on 1 degree of freedom (known as the Mantel-Haenszel or log-rank test statistic)

The log-rank test

- Method (cont.):
  - the log-rank test calculation also produces, for each group, the observed to expected ratio which relates the number of events observed during the follow up with the expected number on the null hypothesis that the survival curve for that group would be the same as that for the combined data
  - see handout from Simpson (in Kerr, Taylor, and Heard)




```
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
```

Roadmap

- Background
- Log-rank test
- Breslow's test
- Others

Breslow's test

- Applicable to data where there is progressive censoring
- More powerful than the log-rank test when the hazard functions are not parallel
- Low power when censoring is high
- Gives more weight to early failures

```
survdiff(Surv(stop, status) ~ clinic, data = dat, na.action =
na.omit, rho = 1)

Call:
survdiff(formula = Surv(stop, status) ~ clinic, data = dat,
na.action = na.omit, rho = 1)

 N Observed Expected (O-E)^2/E (O-E)^2/V
clinic=1 163     77.4     61.4      4.17      16.1
clinic=2  75     19.8     35.8      7.16      16.1

 Chisq= 16.1  on 1 degrees of freedom, p= 6.07e-05
```

Roadmap

- Background
- Log-rank test
- Breslow's test
- Others

Others

- Cox Mantel test
  - similar to the log-rank test
  - applicable to data where there is progressive censoring
  - more powerful than Breslow’s test
- Peto and Peto modification of the Gehan-Wilcoxon test
  - similar to Breslow's test
  - used where the hazard ratio between groups is not constant
- Cox's F test
  - more powerful than Breslow's test if sample sizes are small (n1n2 < 50)

Roadmap

- Background
- Log-rank test
- Breslow's test
- Others