

3 The Log-Rank Test

- We look at 2 groups → extensions to several groups possible
- When are two KM curves statistically equivalent?
 - testing procedure compares the two curves
 - we don't have evidence to indicate that the true survival curves are different
- Nullhypothesis
 - H_0 : no difference between (true) survival curves
- Goal: To find an expression (depending on the data) from which we know the distribution (or at least approximately) under the nullhypothesis

Derivation of test statistic

Remission data: n=42

$t_{(j)}$	# failures		# in risk set	
	m_{1j}	m_{2j}	n_{1j}	n_{2j}
1	0	2	21	21
2	0	2	21	19
3	0	1	21	17
4	0	2	21	16
5	0	2	21	14
6	3	0	21	12
7	1	0	17	12
8	0	4	16	12
10	1	0	15	8
11	0	2	13	8
12	0	2	12	6
13	1	0	12	4
15	0	1	11	4
16	1	0	11	3
17	0	1	10	3
22	1	1	7	2
23	1	1	6	1

Expected cell counts:

$$e_{1j} = \left(\frac{n_{1j}}{n_{1j} + n_{2j}} \right) \times (m_{1j} + m_{2j})$$

↑
↑
 Proportion in risk set # of failures over both groups

$$e_{2j} = \left(\frac{n_{2j}}{n_{1j} + n_{2j}} \right) \times (m_{1j} + m_{2j})$$

EXAMPLE

Expanded Table (Remission Data)

j	$t_{(j)}$	# failures		# in risk set		# expected		Observed-expected	
		m_{1j}	m_{2j}	n_{1j}	n_{2j}	e_{1j}	e_{2j}	$m_{1j} - e_{1j}$	$m_{2j} - e_{2j}$
1	1	0	2	21	21	$(21/42) \times 2$	$(21/42) \times 2$	-1.00	1.00
2	2	0	2	21	19	$(21/40) \times 2$	$(19/40) \times 2$	-1.05	1.05
3	3	0	1	21	17	$(21/38) \times 1$	$(17/38) \times 1$	-0.55	0.55
4	4	0	2	21	16	$(21/37) \times 2$	$(16/37) \times 2$	-1.14	1.14
5	5	0	2	21	14	$(21/35) \times 2$	$(14/35) \times 2$	-1.20	1.20
6	6	3	0	21	12	$(21/33) \times 3$	$(12/33) \times 3$	1.09	-1.09
7	7	1	0	17	12	$(17/29) \times 1$	$(12/29) \times 1$	0.41	-0.41
8	8	0	4	16	12	$(16/28) \times 4$	$(12/28) \times 4$	-2.29	2.29
9	10	1	0	15	8	$(15/23) \times 1$	$(8/23) \times 1$	0.35	-0.35
10	11	0	2	13	8	$(13/21) \times 2$	$(8/21) \times 2$	-1.24	1.24
11	12	0	2	12	6	$(12/18) \times 2$	$(6/18) \times 2$	-1.33	1.33
12	13	1	0	12	4	$(12/16) \times 1$	$(4/16) \times 1$	0.25	-0.25
13	15	0	1	11	4	$(11/15) \times 1$	$(4/15) \times 1$	-0.73	0.73
14	16	1	0	11	3	$(11/14) \times 1$	$(3/14) \times 1$	0.21	-0.21
15	17	0	1	10	3	$(10/13) \times 1$	$(3/13) \times 1$	-0.77	0.77
16	22	1	1	7	2	$(7/9) \times 2$	$(2/9) \times 2$	-0.56	0.56
17	23	1	1	6	1	$(6/7) \times 2$	$(1/7) \times 2$	-0.71	0.71
Totals		9	21			19.26	10.74	-10.26	10.26

$$O_i - E_i = \sum_{j=1}^{\# \text{ failure times}} (m_{ij} - e_{ij})$$

$$O_1 - E_1 = -10.26$$

$$O_2 - E_2 = 10.26$$

$$\text{Log-rank statistic} = \frac{(O_2 - E_2)^2}{\text{Var}(O_2 - E_2)}$$

Remark: We could also work with $O_1 - E_1$ and would get the same statistic! Why?

Distribution of log-rank statistic

H_0 : no difference between survival curves

$$\text{Log-rank statistic for two groups} = \frac{(O_2 - E_2)^2}{\text{Var}(O_2 - E_2)} \sim \chi_1^2$$

Idea of the Proof:

- If X is standard normal distributed then X^2 has a χ^2 distribution with 1 df (assuming X to be one-dim)
- Set $X = \frac{O_2 - E_2}{\sqrt{\text{Var}(O_2 - E_2)}}$
- Then X is standardized and appr. normal distributed for large samples
- Hence X^2 , which is exactly our statistic, has appr. a χ^2 distribution.

The Log-Rank Test for Several Groups

- H_0 : All survival curves are the same
- Log-rank statistics for > 2 groups involves variances and covariances of $O_i - E_i$
- $G (\geq 2)$ groups:
log-rank statistic $\sim \chi^2$ with $G - 1$ df

Remarks

■ Alternatives to the Log-Rank Test

Wilcoxon

Tarone-Ware

Peto

Flemington-Harrington



Variations of the log rank test, derived by applying different weights at the j^{th} failure time

Weighting the Test statistic:

$$\frac{\left(\sum_j w(t_j)(m_{ij} - e_{ij}) \right)^2}{\text{Var} \left(\sum_j w(t_j)(m_{ij} - e_{ij}) \right)}$$

Weight at j^{th} failure time

Remarks

- Choosing a Test
 - Results of different weightings usually lead to similar conclusions
 - The best choice is test with most power
 - There may be a clinical reason to choose a particular weighting
 - Choice of weighting should be a priori! Not fish for a desired p-value!

Stratified log rank test

- Variation of log rank test
- Allows controlling for additional („stratified“) variable
- Split data into stratas, depending on value of stratified variable
- Calculate $O - E$ scores within strata
- Sum $O - E$ across strata

Stratified log rank test - Example

- Remission data
- Stratified variable: 3-level variable (LWBC3) indicating low, medium, or high log white blood cell count (coded 1, 2, and 3, respectively)

->lwbc3 = 1

rx	Events observed	Events expected
0	0	2.91
1	4	1.09
Total	4	4.00

->lwbc3 = 2

rx	Events observed	Events expected
0	5	7.36
1	5	2.64
Total	10	10.00

->lwbc3 = 3

rx	Events observed	Events expected
0	4	6.11
1	12	9.89
Total	16	16.00

-> Total

rx	Events observed	Events expected (°)
0	9	16.38
1	21	13.62
Total	30	30.00

(°) sum over calculations within lwbc3

chi2 (1) = 10.14, Pr > chi2 = 0.0014

Treated Group: rx = 0

Placebo Group: rx = 1

Recall: Non-stratified test $\rightarrow \chi^2$ -value of 16.79 and corresponding p-value rounded to 0.0000

Stratified Log-Rank Test for Remission data

■ R-code

```
> data <- read.table("http://www.sph.emory.edu/~dkleinb/surv2datasets/anderson.dat")
> lwbc3 <-
c(1,1,1,2,1,2,2,1,1,1,3,2,2,2,2,2,3,3,2,3,3,1,2,2,1,1,3,3,1,3,3,2,3,3,3,3,2,3,3,3,2,3)
> fit <- survdiff(Surv(data$V1,data$V2)~data$V5+strata(lwbc3))
```

■ Result

```
> fit
Call:
survdiff(formula = Surv(data$V1, data$V2) ~ data$V5 + strata(lwbc3))

              N Observed Expected (O-E)^2/E (O-E)^2/V
data$V5=0  21         9      16.4      3.33      10.1
data$V5=1  21        21      13.6      4.00      10.1

Chisq = 10.1 on 1 degrees of freedom, p = 0.00145
```

Stratified vs. unstratified approach

Log rank unstratified

$$O_i - E_i = \sum_j (m_{ij} - e_{ij})$$

i = group #, j = j th failure time

Log rank stratified

$$O_i - E_i = \sum_s \sum_j (m_{ijs} - e_{ijs})$$

i = group #, j = j th failure time,
 s = stratum #

Stratified or unstratified (G groups)

Under H_0 :

log rank statistic $\sim \chi^2$ with
 $G - 1$ df

Limitation: Sample size may be small within strata

In next chapter: controlling for other explanatory variables!

4 Cox Proportional Hazards Model



- The formula for the Cox PH model is

$$h(t, \mathbf{X}) = h_0(t) \exp\left(\sum_{i=1}^p \beta_i X_i\right)$$

where

$$\mathbf{X} = (X_1, X_2, \dots, X_p)$$

are the explanatory/predictor variables.

Explanation of the Formula

$$h(t, \mathbf{X}) = h_0(t) \exp\left(\sum_{i=1}^p \beta_i X_i\right)$$

- Product of two quantities:
 - $h_0(t)$ is called the baseline hazard
 - Exponential of the sum of β_i and X_i
- X 's zero (no X 's): reduces to baseline hazard
- Baseline hazard is an unspecified function
 - Semi-parametric model
 - Reason for Cox model being popular

Important Properties of the Cox PH Formula

$$h(t, \mathbf{X}) = h_0(t) \exp\left(\sum_{i=1}^p \beta_i X_i\right)$$

- The baseline hazard $h_0(t)$ does not depend on \mathbf{X} but only on t .
- The exponential involves the X 's but not t .
- The X are time-independent
- Proportional Hazard assumption follows

Time Independent Variables

- Not changing over time
 - Example: sex
- Values are set at time $t = 0$
- Variables unlikely to change are often considered time independent
 - Example: smoking status
- Also other variables are sometimes treated as time independent
 - Examples: age, weight

Example: Data

- T = weeks until going out of remission
- X_1 = group status
- X_2 = log WBC (confounder, effect modifier)
- Interaction?
- $X_3 = X_1 \times X_2$ = group status \times log WBC

Same dataset for each model
 $n = 42$ subjects
 T = time (weeks) until out of remission

Model 1: Rx only

Model 2: Rx and log WBC

Model 3: Rx , log WBC, and
 $Rx \times \log WBC$

EXAMPLE

Leukemia Remission Data

Group 1 ($n = 21$)		Group 2 ($n = 21$)	
t (weeks)	log WBC	t (weeks)	log WBC
6	2.31	1	2.80
6	4.06	1	5.00
6	3.28	2	4.91
7	4.43	2	4.48
10	2.96	3	4.01
13	2.88	4	4.36
16	3.60	4	2.42
22	2.32	5	3.49
23	2.57	5	3.97
6+	3.20	8	3.52
9+	2.80	8	3.05
10+	2.70	8	2.32
11+	2.60	8	3.26
17+	2.16	11	3.49
19+	2.05	11	2.12
20+	2.01	12	1.50
25+	1.78	12	3.06
32+	2.20	15	2.30
32+	2.53	17	2.95
34+	1.47	22	2.73
35+	1.45	23	1.97

+ denotes censored observation

Example: R Output Model 1

Call:

```
coxph(formula = surv(time, event) ~ Rx, data = Data, method = "breslow")
```

n= 42

	coef	exp(coef)	se(coef)	z	Pr(> z)
Rx	1.5092	4.5231	0.4096	3.685	0.000229 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

	exp(coef)	exp(-coef)	lower .95	upper .95
Rx	4.523	0.2211	2.027	10.09

Rsquare= 0.304 (max possible= 0.989)

Likelihood ratio test = 15.21 on 1 df, p=9.615e-05

Wald test = 13.58 on 1 df, p=0.0002288

Score (logrank) test = 15.93 on 1 df, p=6.571e-05

```
> model1$loglik[2]
```

```
[1] -86.37962
```

Example: R Output Model 3

Call:

```
coxph(formula = Surv(time, event) ~ RX * logWBC, data = Data, method = "breslow")
```

n= 42

	coef	exp(coef)	se(coef)	z	Pr(> z)
RX	2.3549	10.5375	1.6810	1.401	0.161
logWBC	1.8028	6.0665	0.4467	4.036	5.45e-05 ***
RX:logWBC	-0.3422	0.7102	0.5197	-0.658	0.510

	exp(coef)	exp(-coef)	lower .95	upper .95
RX	10.5375	0.0949	0.3907	284.201
logWBC	6.0665	0.1648	2.5275	14.561
RX:logWBC	0.7102	1.4080	0.2564	1.967

Rsquare= 0.648 (max possible= 0.989)
 Likelihood ratio test= 43.8 on 3 df, p=1.633e-09
 Wald test = 30.6 on 3 df, p=1.030e-06
 Score (logrank) test = 45.9 on 3 df, p=5.95e-10

```
> model3$loglik[2]
```

```
[1] -72.06572
```

$P = 0.510: \frac{-0.342}{-0.520} = -0.66 = Z$ Wald statistic

LR statistic: uses Log likelihood = -72.066

$-2 \ln L$ (log likelihood statistic) = $-2 \times (-72.066)$
 = 144.132

LR (interaction in model 3)

$= -2 \ln L_{\text{model 2}} - (-2 \ln L_{\text{model 3}})$
 $= (-2 \times -72.280) - (-2 \times -72.066)$
 = 144.560 - 144.132 = 0.428

(LR is χ^2 with 1 d.f. under H_0 :
 no interaction.)

0.40 < P < 0.50, not significant

Wald test P = 0.510

Example: R Output Model 2

Call:

```
coxph(formula = Surv(time, event) ~ Rx + logWBC, data = Data, method = "breslow")
```

n= 42

	coef	exp(coef)	se(coef)	z	Pr(> z)
Rx	1.2941	3.6476	0.4221	3.066	0.00217 **
logWBC	1.6043	4.9746	0.3293	4.872	1.11e-06 ***

	exp(coef)	exp(-coef)	lower .95	upper .95
Rx	3.648	0.2742	1.595	8.343
logWBC	4.975	0.2010	2.609	9.486

Rsquare= 0.644 (max possible= 0.989)
 Likelihood ratio test= 43.41 on 2 df, p=3.744e-10
 Wald test = 31.78 on 2 df, p=1.254e-07
 Score (logrank) test = 42.94 on 2 df, p=4.743e-10

```
> model2$loglik[2]
[1] -72.27926
```

Point estimate:

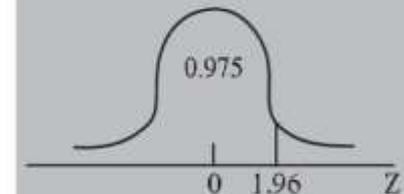
$$\widehat{HR} = 3.648 = e^{1.294}$$

Coefficient of treatment variable

95% confidence interval for the HR:
 (1.595, 8.343)



95% CI for β_1 : $1.294 \pm (1.96)(0.422)$



95% CI for $HR = e^{\beta_1}$:

$$\exp[\hat{\beta}_1 \pm 1.96s_{\hat{\beta}_1}] = e^{1.294 \pm 1.96(0.422)}$$

Example: Continued

Reasons to include logWBC in the model



Confounding: crude versus adjusted HR are meaningfully different
 → must control for logWBC



Precision of confidence intervals: even if no confounding we might prefer to keep logWBC if CI is smaller

Model 1:

	Coef.	Std. Err.	p > z	Haz. Ratio
Rx	1.509	0.410	0.000	4.523

No. of subjects = 42 Log likelihood = -86.380

Model 2:

	Coef.	Std. Err.	p > z	Haz. Ratio
Rx	1.294	0.422	0.002	3.648
log WBC	1.604	0.329	0.000	4.975

No. of subjects = 42 Log likelihood = -72.280

	[95% Conf. Interval]	
Rx model 1	2.027	10.094
	width = 8.067	
Rx model 2	1.595	8.343
log WBC	2.609	9.486

4.2 Why is the Cox PH model popular?

Reasons for the Popularity of the Model

- Robustness
 - Cox model is a “safe” choice of a model in many situations
- Because of the model form:

$$h(t, \mathbf{X}) = \underbrace{h_0(t)}_{\geq 0} \times \underbrace{\exp\left(\sum_{i=1}^p \beta_i X_i\right)}_{\geq 0}$$

the estimated hazards are always non-negative.

- Even though $h_0(t)$ is unspecified we can estimate β_i 's and thus compute the hazard ratio.

Reasons for the Popularity of the Model

- $h(t, \mathbf{X})$ and $S(t, \mathbf{X})$ can be estimated for a Cox model using a minimum of assumptions.
- In survival analysis the Cox model is preferred to a logistic model, since the latter one ignores survival times and censoring information.

4.3 Computing the Hazards Ratio

Definition of the Hazard Ratio

- The Hazard Ratio is defined as

$$HR = \frac{\hat{h}(t, \mathbf{X}^*)}{\hat{h}(t, \mathbf{X})}$$

where

$$\mathbf{X}^* = (X_1^*, X_2^*, \dots, X_p^*)$$

and

$$\mathbf{X} = (X_1, X_2, \dots, X_p)$$

Interpretation of the Hazard Ratio

- Hazard for one individual divided by the hazard for a different individual
- For sake of interpretation we usually want $HR \geq 1$ i.e.

$$\hat{h}(t, \mathbf{X}^*) \geq \hat{h}(t, \mathbf{X})$$

- We thus typically take
 - \mathbf{X}^* : group with larger hazard (e.g. placebo group)
 - \mathbf{X} : group with smaller hazard (e.g. treatment group)

Simplification of the Hazard Ratio

- Baseline hazard cancels out

$$HR = \frac{\hat{h}(t, \mathbf{X}^*)}{\hat{h}(t, \mathbf{X})} = \frac{\hat{h}_0(t) \exp\left(\sum_{i=1}^p \hat{\beta}_i X_i^*\right)}{\hat{h}_0(t) \exp\left(\sum_{i=1}^p \hat{\beta}_i X_i\right)} = \exp\left(\sum_{i=1}^p \hat{\beta}_i (X_i^* - X_i)\right)$$

Example: Remission Data, Model 1

- Only one variable of interest: exposure status
 - Placebo group: $X_1^* = 1$
 - Treatment group: $X_1 = 0$
- Hazard Ratio simplifies to

$$HR = \exp\left(\hat{\beta}_1 (X_1^* - X_1)\right) = e^{\hat{\beta}_1}$$

- Since $\hat{\beta}_1 = 1.509$

we have $HR = 4.523$

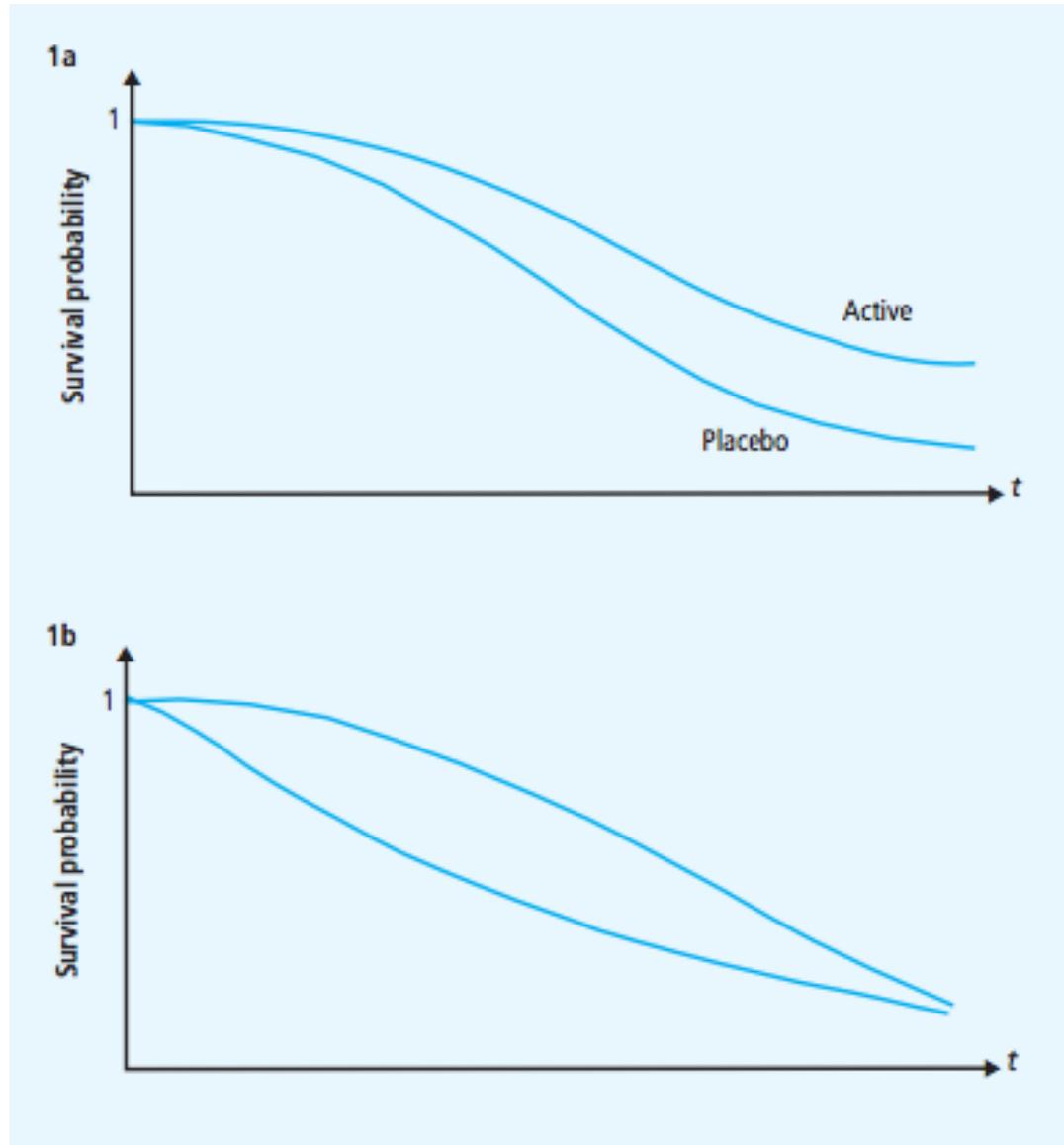
4.4 Meaning of the Proportional Hazards Assumption

Meaning of the PH Assumption

- Remember that the PH assumption requires that the HR is constant over time

$$HR = \exp \left(\sum_{i=1}^p \hat{\beta}_i (X_i^* - X_i) \right)$$

Which one violates the assumption?





The End!
Thank you!