
SISCER Mod 12

Survival Analysis

Lecture 3

Ying Qing Chen, Ph.D.

Department of Medicine
Stanford University

Department of Biostatistics
University of Washington

In Lecture 2

- What we discussed
 - Kaplan-Meier curves/estimates
 - Nonparametric MLE



Two-sample testing

- Two-sample studies
 - Compare two survival functions via hypothesis testing
 - Assess group differences or treatment effect
 - Superiority versus non-inferiority
 - Active-control
 - Sample size and power calculation
 - Sequential monitoring
 - Establish efficacy/futility boundaries
 - Project timing of interim analyses
-

Review of binary outcomes

- In a two by two table

		D	\bar{D}	
Treatment	A	a	b	n_A
	B	c	d	n_B
		m_D	$m_{\bar{D}}$	n

- Chi-square

Null hypothesis $H_0 : p_A = p_B$

$$T = \left(\frac{a - m_D \left(\frac{n_A}{n} \right)}{\sqrt{\frac{n_A n_B m_D m_{\bar{D}}}{n^2(n-1)}}} \right)^2$$

when n is large, $T \sim \chi^2(1)$.

A simple two-sample test

Suppose t -year survival rate is of interest

$$H_0 : S_A(t) = S_B(t).$$

Data could be censored before t . We use the K-M estimate to estimate $S_A(t)$ and $S_B(t)$, and construct a test statistic

$$T = \frac{\hat{S}_A(t) - \hat{S}_B(t)}{\widehat{SD}[\hat{S}_A(t) - \hat{S}_B(t)]} \sim N(0, 1).$$

Here $SD[\hat{S}_A(t) - \hat{S}_B(t)]$ can be estimated by Greenwood's formula,

$$\text{Var}[\hat{S}_A(t) - \hat{S}_B(t)] = \text{Var}(\hat{S}_A(t)) + \text{Var}(\hat{S}_B(t))$$

$$\widehat{SD}[\hat{S}_A(t) - \hat{S}_B(t)] = \sqrt{\widehat{\text{Var}}(\hat{S}_A(t)) + \widehat{\text{Var}}(\hat{S}_B(t))},$$

where $\widehat{\text{Var}}$ is derived by by Greenwood's formula.

■ Shortfalls

- ❑ Only test survival difference at a specific time
 - ❑ Yet to provide a overall summary of difference over time
 - ❑ Yet to take advantage of time-varying incidence rates to maximize the power to detect difference
-

Log-rank test

■ Motivation

- Synthesize two-by-two tables that can be constructed at each observed failure time
 - Why not censoring?
- For multiple two-by-two tables
 - Mantel-Haenszel approach when adjusting for potential confounders by stratification



Log-rank test

- Ideas:
1. Create a 2×2 table at each uncensored failure time
 2. The construction of each 2×2 table is based on the corresponding risk set.
 3. Combine information from tables

At an uncensored time

		D	\bar{D}	
Treatment	A	d	$n_A - d$	n_A
Treatment	B	$m_D - d$	$n_B - (m_D - d)$	n_B
		m_D	$m_{\bar{D}}$	N

N : # individuals in the risk set at y from pooled data

d : # failures at y from group A

m_D : # failures at y from pooled data

n_A : # individuals in the risk set at y from group A

n_B : # individuals in the risk set at y from group B

$m_{\bar{D}} = N - m_D$

Log-rank test

■ Test statistic:

$$Z = \frac{\sum_{i=1}^k (D_{(i)} - E_0[D_{(i)}])}{\sqrt{\sum_{i=1}^k \text{Var}_0(D_{(i)})}} \underset{n \text{ large}}{\sim} N(0, 1)$$

- Key quantities

- ▷ Observed: $O_i = d_{1i}$

- ▷ Expected: $E_i = n_{1i} \cdot \frac{D_i}{N_i}$, or $E_i = D_i \cdot \frac{n_{1i}}{N_i}$

- ▷ Variance: $V_i = n_{1i}n_{2i}D_iS_i/[N_i^2(N_i - 1)]$ (hypergeometric)

- Statistic:

$$X_L^2 = \frac{\left[\sum_{i=1}^J (O_i - E_i) \right]^2}{\sum_{i=1}^J V_i}$$

- Under the null hypothesis $X_L^2 \sim \chi^2(\text{df} = 1)$

Underlying theory

Use the following method to construct the test statistic:
conditional on $n_A, n_B, m_D, m_{\bar{D}}$, the random number d follows a hypergeometric distribution (under H_0) with probability

$$\frac{\binom{n_A}{d} \binom{n_B}{m_D - d}}{\binom{N}{m_D}}, \quad \max(0, m_D - n_B) \leq d \leq \min(n_A, m_D).$$

Under H_0 ,

$$E_0(D) = m_D \left(\frac{n_A}{N} \right)$$

$$\text{Var}_0(D) = \frac{n_A n_B m_D m_{\bar{D}}}{N^2 (N - 1)}$$

Two-sided testing procedure

Usually the significance level of a test is set up to be 0.05.

use

$$Z^2 = \left[\frac{\sum_1^k (D_{(i)} - E_0[D_{(i)}])}{\sqrt{\sum_1^k \text{Var}_0(D_{(i)})}} \right]^2 \underset{n \text{ large}}{\sim} \chi^2(1)$$

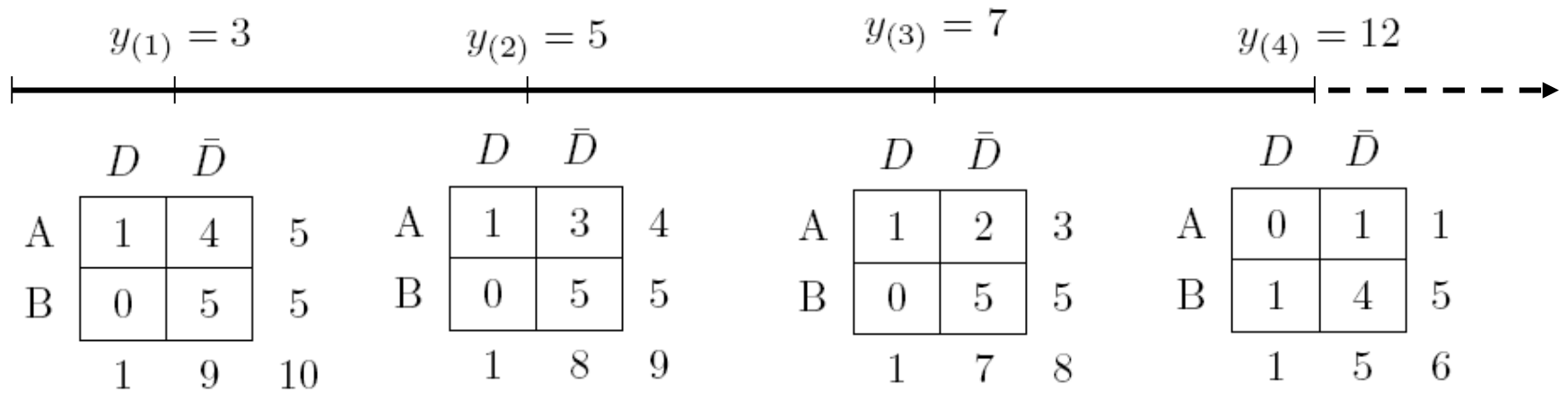
Reject H_0 when $z^2 > 3.84$ ($|z| > 1.96$)

p -value = Probability for values larger than z^2 .

Example

Example. Group A 3, 5, 7, 9⁺, 18
Group B 12, 19, 20, 20⁺, 33⁺

Uncesored: 3, 5, 7, 12, 18, 19, 20



$y_{(i)}$	$d_{(i)}$	$E_0[d_{(i)}]$	$\text{Var}_0[d_{(i)}]$
3	1	$1 \times \frac{5}{10} = 0.5$	$\frac{5 \times 5 \times 1 \times 9}{10^{2.9}} = 0.25$
5	1	$1 \times \frac{4}{9} = 0.44$	$\frac{4 \times 5 \times 1 \times 8}{9^{2.8}} = 0.2469$
7	1	$1 \times \frac{3}{8} = 0.38$	0.2344
12	0	$1 \times \frac{1}{6} = 0.17$	0.1389
18	1	$1 \times \frac{1}{5} = 0.20$	0.1600
19	0	$1 \times \frac{0}{4} = 0$	0
20	0	$1 \times \frac{0}{3} = 0$	0

$$\sum_1^7 (d_{(i)} - E_0(d_{(i)})) = (1 - 0.5) + \dots + (0 - 0) = \underline{2.31}$$

$$\sum_1^7 \text{Var}_0(d_{(i)}) = 0.25 + \dots + 0 = \underline{1.030}$$

$$z = \frac{2.31}{\sqrt{1.030}} = \underline{2.28} \quad \longrightarrow \quad z^2 = (2.28)^2 = 5.198 > 3.84$$

$$\longrightarrow \quad p\text{-value} = 0.0226 \Rightarrow \text{reject } H_0$$

Relationship between Mantel-Haenszel and log-rank

- **Mantel-Haenszel:** series of (independent) tables: different levels of a confounder:

	Exposed (E)	Unexposed (\bar{E})
Diseased (D)	a_i	b_i
non-Diseased (\bar{D})	c_i	d_i

- M-H compares $P(D | E, C = i)$ and $P(D | \bar{E}, C = i)$ and is designed for the situation where the odds ratios, Ψ_i are constant across strata.
 - ▷ $H_0 : \Psi_i \equiv 1$ for all i ; $H_1 : \Psi_i \equiv \Psi \neq 1$.
 - ▷ where

$$\Psi_i = \frac{P(D | E, C = i) / P(\bar{D} | E, C = i)}{P(D | \bar{E}, C = i) / P(\bar{D} | \bar{E}, C = i)}$$

- LogRank: series of (dependent) tables: different observed death times:

	Group 1	Group 2	
Deaths at $t_{(i)}$	d_{1i}	d_{2i}	D_i
Survivors at $t_{(i)}$	s_{1i}	s_{2i}	S_i
	n_{1i}	n_{2i}	N_i

- Thus we would expect the LogRank test to be powerful when “odds ratios” over infinitesimal time intervals were constant across time.

- This idea implies $C_t \equiv C$ where

$$C_t = \frac{P(T \in [t, t + \Delta t) | \mathbf{E}, T \geq t) / [1 - P(T \in [t, t + \Delta t) | \mathbf{E}, T \geq t)]}{P(T \in [t, t + \Delta t) | \bar{\mathbf{E}}, T \geq t) / [1 - P(T \in [t, t + \Delta t) | \bar{\mathbf{E}}, T \geq t)]}$$

- As $\Delta t \rightarrow 0$ we have
 1. $1 - P$'s go to 1.0
 2. Ratio of P 's same as ratio of $P/\Delta t$'s
 3. So Ratio of P 's $\rightarrow \frac{\lambda(t|\mathbf{E})}{\lambda(t|\bar{\mathbf{E}})} = RR_t$
- So the LogRank test is designed for (and is most powerful for) the situation where

$$RR_t = \frac{\lambda(t | \mathbf{E})}{\lambda(t | \bar{\mathbf{E}})} = C$$

- Note:

$$\begin{aligned} O_i - E_i &= d_{1i} - n_{1i} \frac{D_i}{N_i} = d_{1i} - D_i \frac{n_{1i}}{N_i} \\ &= \frac{n_{1i}n_{2i}}{N_i} \left(\frac{d_{1i}}{n_{1i}} - \frac{d_{2i}}{n_{2i}} \right) \end{aligned}$$

- This shows that the statistic is based upon a weighted comparison of the estimated hazards (d_{1i}/n_{1i}) and (d_{2i}/n_{2i}) where the weight is $n_{1i}n_{2i}/N_i$.

Weighted log-rank test

- **Q:** What happens when the hazards are not proportional?
 - ▷ Similar to M-H, as long as the difference between the hazards has a consistent sign, the LogRank test usually “does well”.
- Other tests are available. One family has the form:

$$X_W^2 = \frac{\left[\sum_{i=1}^J w_i (O_i - E_i) \right]^2}{\sum_{i=1}^J w_i^2 V_i}$$

where the weights $w_i = w[t_{(i)}]$ can place emphasis on a particular time range (e.g. early, late).

Weighted log-rank test

- **Weight:** Let the weight w_i be a function of the number of subjects at-risk at time $t_{(i)}$, N_i :
 - ▷ $w_i = 1 \Rightarrow$ LogRank Test
(Mantel, 1966).
 - ▷ $w_i = N_i \Rightarrow$ Wilcoxon-Gehan-Breslow Test
(Gehan 1965; Breslow 1970).
 - ▷ $w_i = \sqrt{N_i} \Rightarrow$ Tarone-Ware Test
(Tarone and Ware 1977).
- The LogRank emphasizes the tail of the survival (relatively).
- The Wilcoxon-Gehan-Breslow emphasizes the beginning of the curve.

-
- **Q:** Which to choose?
 - ▷ Which is scientifically more important – early versus late?
 - ▷ The LogRank is most powerful for proportional hazards, so naturally corresponds to Cox regression.
 - There are additional tests such as those suggested by Harrington and Fleming (1982) that use $w_i = [\hat{S}(t_{(i-1)})]^\rho$ for some value of ρ . This will weight earlier times (depending on the choice of ρ) but has an advantage in that it depends less on the censoring distribution than the Wilcoxon test.
 - $\rho = 1$ using $w_i = \hat{S}(t_{(i-1)})$ is the **Peto-Prentice** generalization of the Wilcoxon.
-

-
- The use of normality (chi-square) for X_L^2 stems from having a large number of (possibly sparse) tables over which we sum – that is, the key assumption is that J , the number of observed event times is large. This is the same as M-H where M-H could be validly applied to matched data (e.g. each 2×2 is only a pair of subjects).
 - Similar to M-H normality could be assumed for a few large tables – this would imply heavily “tied” data (e.g. 2,2,2, 7,7,7,7 for outcome).
 - Exact methods have been proposed when groups are not of equal size and one group is small (see Heinze, Gnant, and Schemper, 2003).
-

A trivia

- **Q:** Why is this named “log rank”?
- **A:** Kalbflesich and Prentice (2002) p. 27:

“The name log-rank was coined by Peto and Peto (1972) and the motivation of the term is not entirely clear to all – some say to apply it one first logs the data and then ranks them.”

Take-home message

■ Log-rank test is essentially to test

- **Hypotheses**

- ▷ $H_0: S_1(t) = S_2(t)$ for all t .

- ▷ $H_1: S_1(t) = [S_2(t)]^C$, for $C \neq 1$

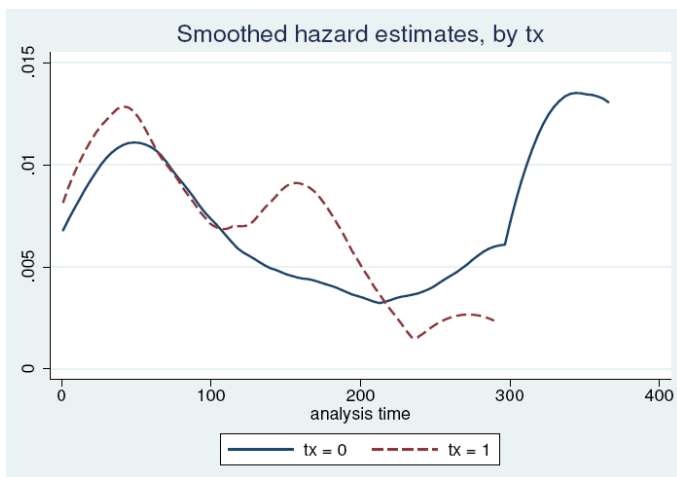
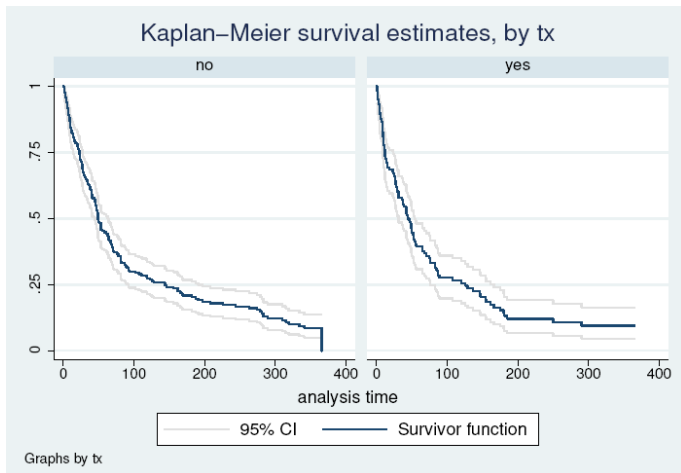
- **Hypotheses** (equivalently)

- ▷ $H_0: \lambda_1(t) = \lambda_2(t)$ for all t .

- ▷ $H_1: \lambda_1(t) = C \cdot \lambda_2(t)$, for $C \neq 1$

Proportional hazards

An example



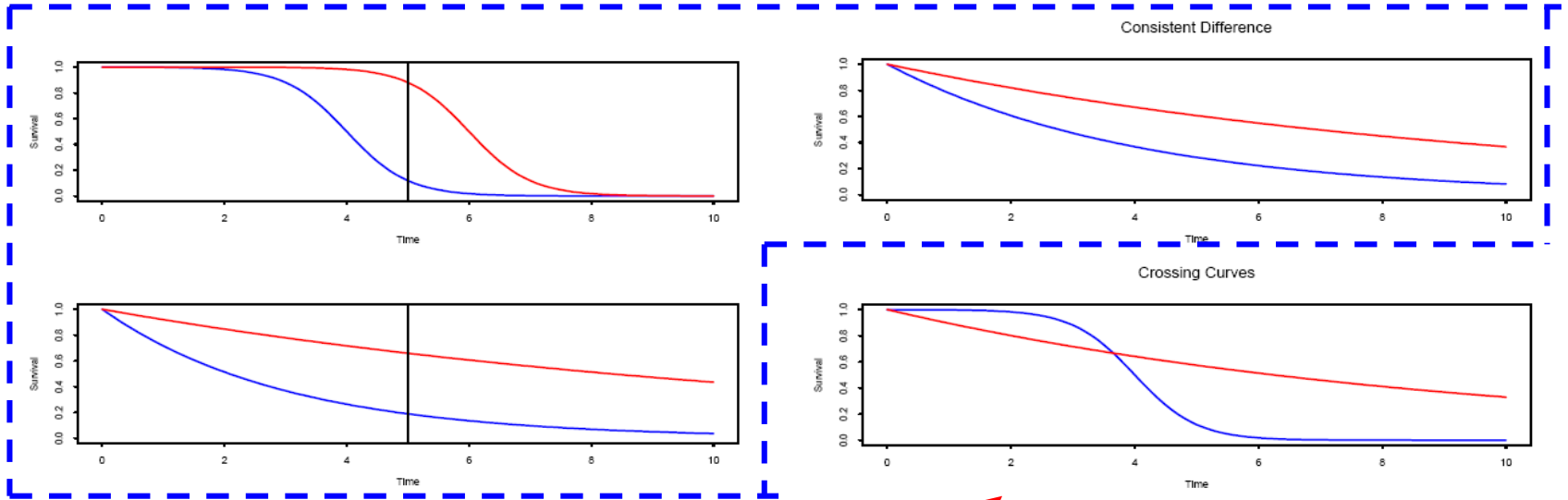
Log-rank test for equality of survivor functions

tx	Events observed	Events expected
no	180	187.42
yes	103	95.58
Total	283	283.00

$\chi^2(1) = 0.89$
 $\text{Pr} > \chi^2 = 0.3465$

→ **Not significant**

Hypothetical examples



Yes?

No?

Log-rank test

Stratified log-rank

- **Q:** What if we want to test for differences in risk, adjusting for some potential confounding factor?
- **Solution:** Stratify on (categorical) confounder.
- Suppose stratification factor has K levels: $k = 1, 2, \dots, K$.
- Test
 - ▷ $H_0 : \lambda_k(t | \mathbf{E}) \equiv \lambda_k(t | \bar{\mathbf{E}})$
for all t and $k = 1, \dots, K$.
 - ▷ $H_1 : \lambda_k(t | \mathbf{E}) = C \cdot \lambda_k(t | \bar{\mathbf{E}})$
for all t and $k = 1, \dots, K$, and $C \neq 1$.

- Test Statistic:

$$X_L^2 = \frac{\left[\underbrace{\sum_{i=1}^{J_1} (O_{1i} - E_{1i})}_{\text{strata 1}} + \underbrace{\sum_{i=1}^{J_2} (O_{2i} - E_{2i})}_{\text{strata 2}} + \dots + \underbrace{\sum_{i=1}^{J_K} (O_{Ki} - E_{Ki})}_{\text{strata K}} \right]^2}{\underbrace{\sum_{i=1}^{J_1} V_{1i}}_{\text{strata 1}} + \underbrace{\sum_{i=1}^{J_2} V_{2i}}_{\text{strata 2}} + \dots + \underbrace{\sum_{i_K=1}^{J_K} V_{Ki}}_{\text{strata K}}}$$

- Where O_{ki} , E_{ki} , and V_{ki} are calculated solely from subjects in strata k .
- Under H_0 we have $X_L^2 \sim \chi^2(\text{df} = 1)$.

STATA codes

- STATA: `sts test tx, strata(group)`

Stratified log-rank test for equality of survivor functions

		Events	Events
tx		observed	expected(*)
no		223	231.14
yes		151	142.86
Total		374	374.00

(*) sum over calculations within group

chi2(1) = 0.79 Pr>chi2 = 0.3754

One-way ANOVA

- Now consider the situation of comparing more than two groups.
- | |
|------------|
| Hypotheses |
|------------|

 - ▷ $H_0 : \lambda_1(t) = \lambda_2(t) = \dots = \lambda_K(t)$
 - ▷ $H_1 : \text{at least one inequality among } K \text{ groups.}$
- | |
|------|
| Data |
|------|

 at each $t_{(i)}$ based on all of the data:

$t_{(i)}$	Group 1	Group 2	...	Group k	...	Group K	
Deaths	d_{1i}	d_{2i}	...	d_{ki}	...	d_{Ki}	D_i
Survivors	s_{1i}	s_{2i}	...	s_{ki}	...	s_{Ki}	S_i
	n_{1i}	n_{2i}	...	n_{ki}	...	n_{Ki}	N_i

- Similar to before, we accumulate the difference between observed for group k and that which would be expected under the null hypothesis.
- The we use a multivariate generalization of $\frac{(O-E)^2}{V}$ to form the test statistic.
- Key quantities for group k :
 - ▷ Observed: $O_{ki} = d_{ki}$
 - ▷ Expected: $E_{ki} = n_{ki} \cdot \frac{D_i}{N_i}$
 - ▷ Variance: $V_{kki} = \frac{n_{ki}(N_i - n_{ki})D_i S_i}{N_i^2(N_i - 1)}$
 - ▷ Covariance with $d_{k'i}$: $V_{kk'i} = \frac{-n_{ki}n_{k'i}D_i S_i}{N_i^2(N_i - 1)}$

- Accumulate:

- ▷ $O_k - E_k = \sum_{i=1}^J (O_{ki} - E_{ki})$

- ▷ $V_{kk} = \sum_{i=1}^J V_{kki}$

- ▷ $V_{kk'} = \sum_{i=1}^J V_{kk'i}$

- Test Statistic

$$X^2 = \begin{pmatrix} O_1 - E_1 \\ O_2 - E_2 \\ \vdots \\ O_K - E_k \end{pmatrix}^T \begin{bmatrix} V_{11} & V_{12} & \dots & V_{1K} \\ V_{21} & V_{22} & & V_{2K} \\ \vdots & & \ddots & \vdots \\ V_{K1} & V_{K2} & & V_{KK} \end{bmatrix} \begin{pmatrix} O_1 - E_1 \\ O_2 - E_2 \\ \vdots \\ O_K - E_k \end{pmatrix}$$

-
- The form of the test statistic is called a “quadratic form”.
 - The matrix notation V^{-} denotes a generalization of $1/V$ used for a single number.
 - Under the null hypothesis
 - ▷ $X^2 \sim \chi^2(\text{df} = K - 1)$
 - Note: the negative covariance $V_{kk'i}$ between d_{ki} and $d_{k'i}$ comes from conditioning on the margin totals, D_i , S_i , n_{ki} , and $n_{k'i}$ – that is, if deaths sum to D_i then allowing an increase in deaths for group k would necessitate a decrease in other groups (possibly k').
-

STATA codes

- STATA: sts test treat

Log-rank test for equality of survivor functions

		Events	Events
treat		observed	expected
-----+-----			
none		223	223.09
topical Acy		30	38.34
oral Acy		111	95.96
IV Acy		10	16.61
-----+-----			
Total		374	374.00
		chi2(3) =	6.89
		Pr>chi2 =	0.0755

Trend test

- In some situations we have K groups that can be ordered based on some scale, or “dose” vector (x_1, x_2, \dots, x_K) .
- In these situations we are interested in testing whether the hazard functions tend to increase or decrease with “dose”.
- | |
|------------|
| Hypotheses |
|------------|

 - ▷ $H_0 : \lambda_1(t) = \lambda_2(t) = \dots = \lambda_K(t)$
 - ▷ $H_1 : \lambda_1(t) \geq \lambda_2(t) \geq \dots \geq \lambda_K(t)$, or
 $\lambda_1(t) \leq \lambda_2(t) \leq \dots \leq \lambda_K(t)$, with at least one $>$ or $<$.
- | |
|------------|
| Hypotheses |
|------------|

 (more specifically)
 - ▷ $H_0 : \lambda_1(t) = \lambda_2(t) = \dots = \lambda_K(t)$
 - ▷ $H_1 : C^{x_1} \cdot \lambda_1(t) = C^{x_2} \cdot \lambda_2(t) = \dots = C^{x_K} \cdot \lambda_K(t), C \neq 1.$

- **Q:** Does risk increase (or decrease) with increasing dose?
- Test Statistic:

$$X_T^2 = \frac{\left[\sum_{k=1}^K x_k \cdot (O_k - E_k) \right]^2}{\sum_{k=1}^K x_k^2 \cdot V_{kk} + 2 \cdot \sum_{k < k'} x_k \cdot x_{k'} \cdot V_{kk'}}$$

- Under null $X_T^2 \sim \chi^2(\text{df} = 1)$.
- Test is effectively for a regression of $\log \lambda(t)$ on x_k , using $\log \lambda(t | x_k) = \alpha(t) + \beta \cdot x_k$.

STATA codes

- STATA sts test group, trend

Log-rank test for equality of survivor functions

	Events	Events
group	observed	expected

I	36	94.86
II	283	245.73
III	55	33.41

Total	374	374.00
-------	-----	--------

chi2(2) = 59.08

Pr>chi2 = 0.0000

Test for trend of survivor functions

chi2(1) = 57.49

Pr>chi2 = 0.0000

Counting process representation of weighted log-rank test

- A general form

$$W = \int_0^{\infty} W(u) \frac{Y_1(u)Y_2(u)}{Y_1(u) + Y_2(u)} \left\{ d\hat{\Lambda}_1(u) - d\hat{\Lambda}_2(u) \right\}$$

- statistics of the class K (Gill, 1980)
 - $W(\cdot)$: weight function
 - weighted differences in cumulative hazard functions
- Choices of weight function $W(\cdot)$
 - $W(t) = 1$: Log-rank
 - $W(t) = c \cdot Y(t)$: Wilcoxon rank-sum, Gehan-Wilcoxon
 - $W(t) = \hat{S}(t-)$: Prentice-wilcoxon
 - $W(t) = \hat{S}(t-)^{\rho}, \rho > 0$: G^{ρ} -family
 - $W(t) = \hat{S}(t-)^{\rho} [1 - \hat{S}(t-)]^{\gamma}$: $G^{\rho\gamma}$ -family
 - $W(t) = K(t) [n_1 n_2 / (n_1 + n_2)]^{1/2} [Y_1(t) + Y_2(t)] / Y(t)$: the class K

Power calculation based on weighted log-rank test

- Power analysis of weighted Log-rank test statistics
 1. type-I error: $\alpha = 5\%$
 2. power level
 3. alternative hypothesis
 4. error bound
-

- Under $H_0 : \lambda_0(t) = \lambda_1(t) = \lambda(t)$
- Alternative hypothesis
 - $H_1 : \lambda_1(t) = \lambda_0(t)e^{\beta_n \times \theta(t)}$
 - $\log[\lambda_1(t | Z_i) / \lambda_0(t)] = \beta_n Z_i \times \theta(t)$
 - $\theta(t)$: take into account of nonproportionality
 - β_n : distance between the null and an alternative
- Given a sample size n ,

$$\text{Power} = \Pr \left\{ \left| n^{-1/2} W / \sqrt{\hat{\alpha}(\tau)} \right| > z_{1-\alpha/2} \mid H_1 \right\}$$

Distributional properties and power

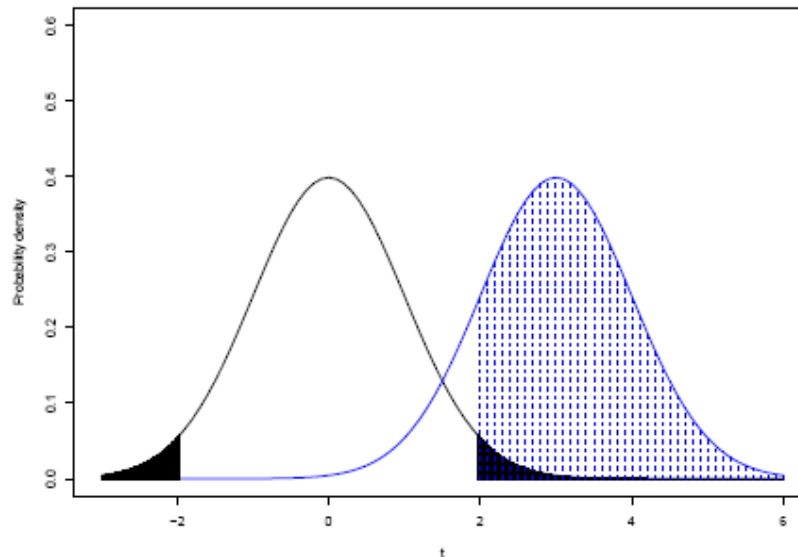
- Summary on $n^{-1/2}W$

- Under H_0 : $n^{-1/2}W \sim \mathcal{N}(0, A(\underline{w}^2))$

Weight function
user-defined

- Under H_{1n} : $n^{-1/2}W \sim \mathcal{N}(\xi A(\underline{\theta}w), A(\underline{w}^2))$

Degree of nonproportionality



$n^{1/2}\beta_n \rightarrow \underline{\xi}$ Difference to detect

$$P = \Phi \left(\frac{\xi A(\theta w)}{A(w^2)^{1/2}} - z_{1-\alpha/2} \right)$$

$$A(w^2) = \int_0^\tau w(u)^2 E_{H_0}[(Z_i - \mu_Z(u))^2 I(X \geq u)] \lambda_0(u) du$$

Sample size calculation

- In practice, we have a fixed β_0 to be detected
 - $H_0 : \lambda_1(t) = \lambda_0(t)$
 - $H_1 : \lambda_1(t) = \lambda_0(t)e^{\beta_0 \times \theta(u)}$
- Standardized weighted Log-rank TS :
 - under H_0 : $TS \sim \mathcal{N}(0, 1)$
 - under H_1 :

$$TS \simeq \mathcal{N} \left(\frac{n^{1/2} \beta_0 A(\theta w)}{A(w^2)^{1/2}}, 1 \right)$$

- Power $P = \Pr\{|TS| \geq z_{1-\alpha/2}\} = 1 - \beta$

–

$$\frac{n^{1/2}\beta_0 A(\theta w)}{A(w^2)^{1/2}} = z_{1-\alpha/2} + z_{1-\beta} \Rightarrow n = \frac{(z_{\alpha/2} + z_{\beta})^2 A(w^2)}{\beta_0 A(\theta w)^2}$$

– $w = \theta = 1$

* Log-rank for proportional hazards model

* sample size

$$n = \frac{(z_{\alpha/2} + z_{\beta})^2}{\beta_0^2 A(1)}$$

– what is $A(1)$?

* recall on $A(1) = \int_0^\infty E[(Z - \mu_Z(u))^2 I(X \geq u)] \lambda_0(u) du$

* $A(1) = \pi_Z(1 - \pi_Z) \Pr(\Delta = 1)$

Proportion of treatment arm
e.g. 1-to-1 randomization: 1/2

– Sample size is then

$$n \Pr(\Delta = 1) = \frac{(z_{\alpha/2} + z_\beta)^2}{\beta_0^2 \pi_Z (1 - \pi_Z)}$$

– Expected # failures/events: $E_D = n \Pr(\Delta = 1)$

* $HR = e^\beta$ is hazards ratio

* 1-to-1 treatment-control assignment

*

$$E_D = \frac{4(z_{\alpha/2} + z_\beta)^2}{(\log HR)^2}$$

– Example:

* type-I error: 5%

* power: 90%

* $HR = 2$

* $E_D = 42 / (\log HR)^2 = 88$

Sequential monitoring

- Develop a design for repeated data analyses
 - satisfying the ethical need for early termination if initial results are extreme
 - not increasing the chance of false conclusions
-

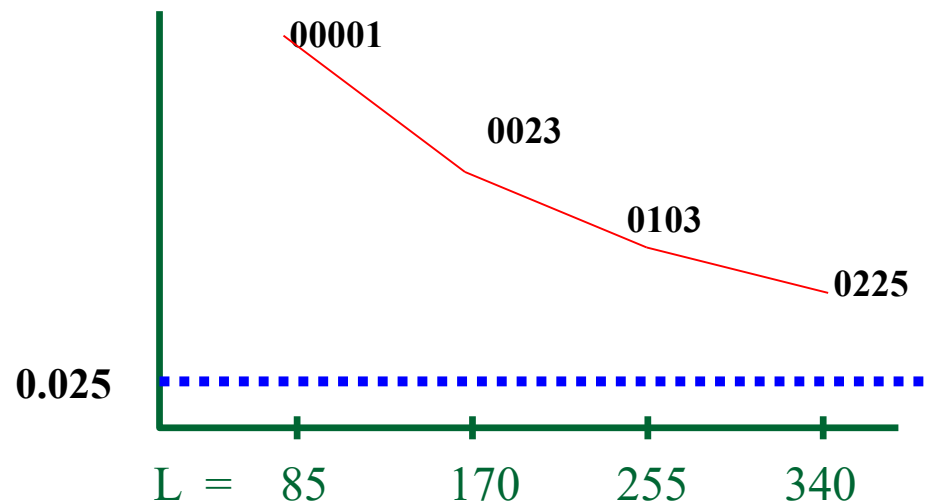
O'Brien-Fleming guideline

- How the O'Brien-Fleming guideline works:
 - Arriving at recommendations about early termination of clinical trials
 - that establish favorable effects
 - that rule out favorable effects
-

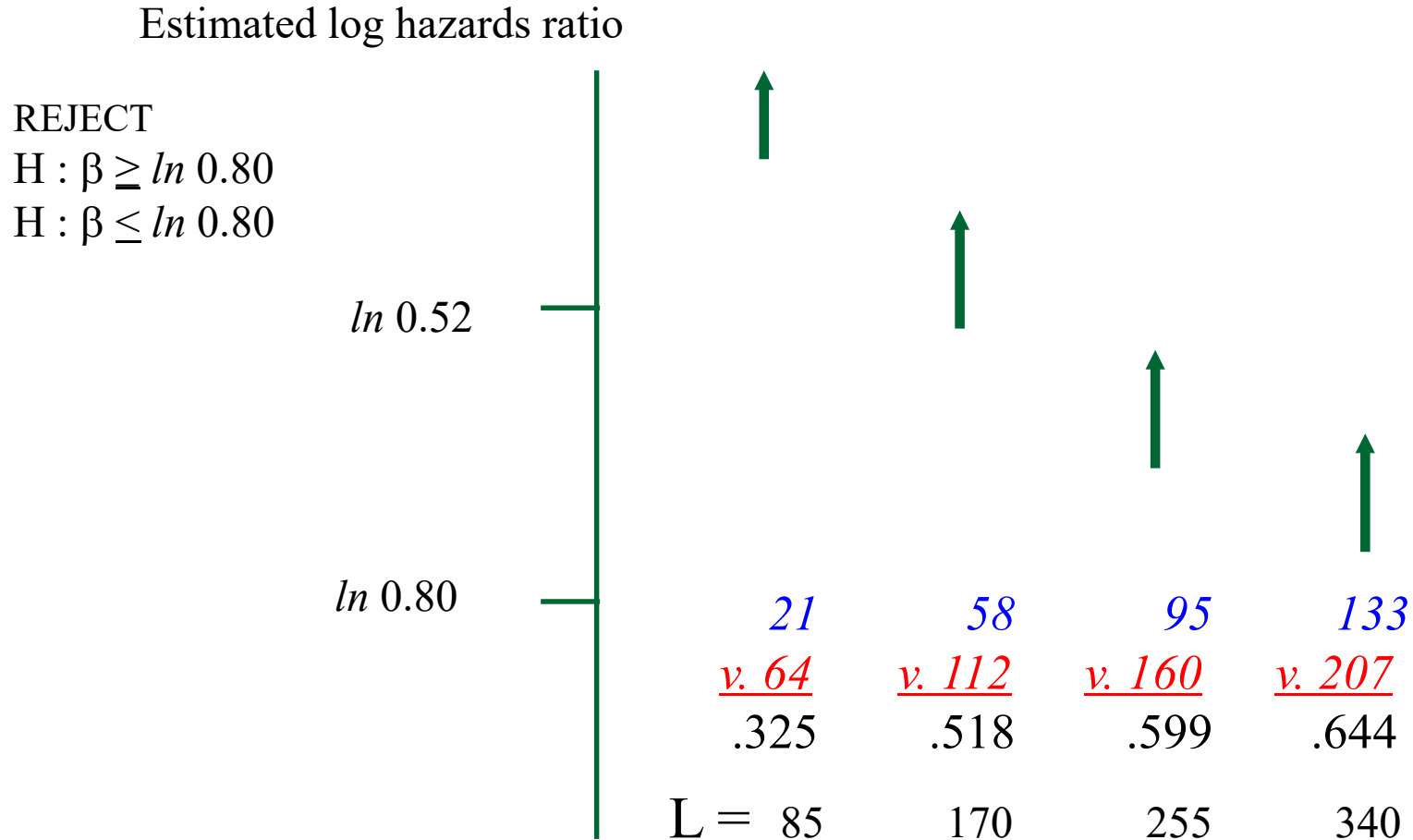
Example: O'Brien-Fleming boundaries

- Example: HPTN 052
 - Goal: with 4 analyses, preserve the 1-sided false positive error rate: 0.025
 - Issue: account for multiple comparisons

O'Brien-Fleming Boundaries, *Biometrics* (1979)



Establish favorable effects



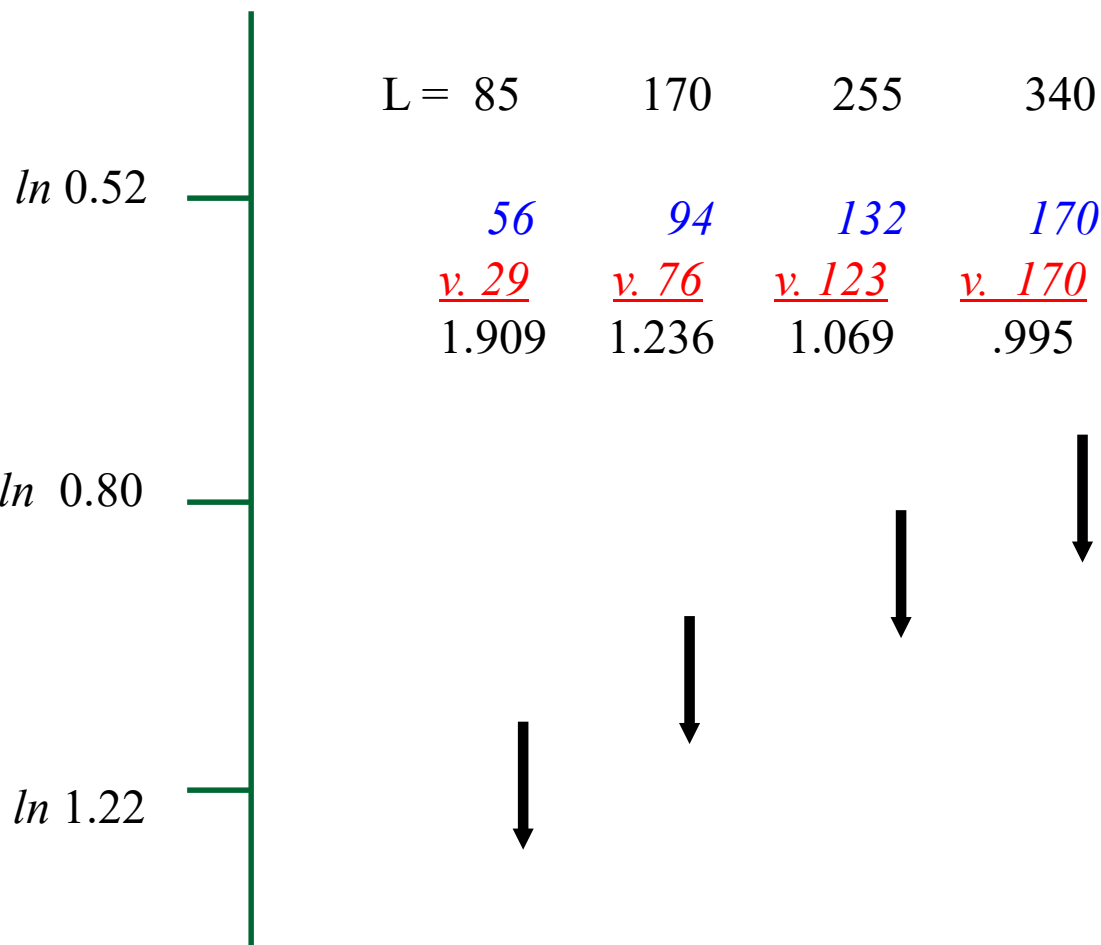
Rule out favorable effects

Estimated log hazards ratio

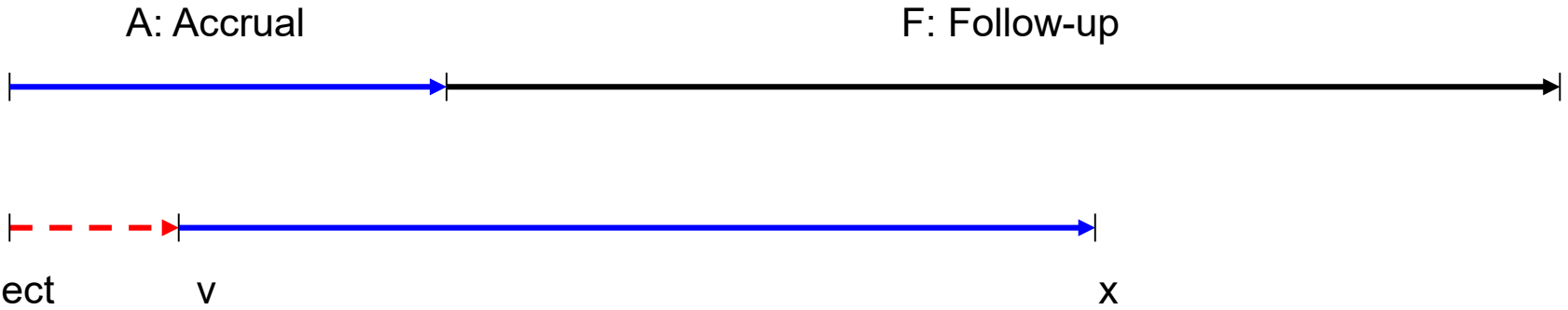
REJECT

$H : \beta \geq \ln 0.80$

$H : \beta \leq \ln 0.80$



Accrual and follow-up



- Step 1: what is the probability of a subject to have an event at x when he/she is recruited at v ?

$$Q(x) = \int_0^x \lambda(u) \exp\{-[\Lambda(u) + H(u)]\} du \implies Q(F - v)$$

where $\Lambda(u) = \int_0^u \lambda(v) dv$ and $H(u) = \int_0^u h(v) dv$

Rate of lost-to-follow-up

-
- Step 2: Total expected number of events

$$N(A, F) = \int_0^A a(v) = Q(F - v)dv$$

Example: HTPN 052 Trial

- Two-arm control randomized clinical trial of treatment strategies of ART management
 - Immediate versus delayed
 - Four interim analyses planned
- $N=1750$
- Accrual time: $A=2.5$ years or 3 years
- Accrual rate: $a(u) = a = (1750/2)/2.5 = 350$ or $(1750/2)/3 = 292$
- Time-to-infection event rate: $\lambda(u)$ as stated in the first section.
 - using high effective rate for the immediate arm.
 - using 50% as reduction rate to generate the event rate for the delayed arm.
- Loss-to-follow-up rate: $h(u) = 6\% = h^*$

Results:

1. When $A=2.5$ years and vary F from 3.5 to 6.5 years

Accrual time (A)	2.5	2.5	2.5	2.5	2.5	2.5	2.5
Follow-up time (F)	3.5	4	4.5	5	5.5	6	6.5
Total time (T)	6	6.5	7	7.5	8	8.5	9
Total events (N_{event})	165	168	173	176	181	184	188

2. when $A=3$ years and vary F from 3.5 to 6.5 years

Accrual time (A)	3	3	3	3	3	3	3
Follow-up time (F)	3.5	4	4.5	5	5.5	6	6.5
Total time (T)	6.5	7	7.5	8	8.5	9	9.5
Total events (N_{event})	166	171	174	179	182	186	189