Syllabus

1. Introduction
   - Survival data
   - Censoring mechanism
   - Application in medical field
2. Concepts and definitions
   - Survival function
   - Hazard function
3. Non-parametric approach
   - Life table
   - Kaplan-Meier survival estimate
   - Hazard function
   - Median and percentile survival time
4. Hypothesis testing
   - Overview – hypothesis, test statistics, p-values
   - Log-rank
   - Wilcoxon
   - Gehan test
5. **Study design and sample size estimation**
   - **Overview**
   - **Survival sample size estimation**
   - **Accrual time and Study duration**
6. Semiparametric model – proportional hazard model
   - Partial likelihood
   - Inference
   - Time varying covariates
   - Stratification
7. Model checking in the PH model
   - Model checking
   - Residuals
8. Parametric model
   - Parametric proportional hazard model
   - Accelerate failure model
9. Other topics
   - Competing risk
   - Recurrent events
   - Non-proportional hazard ratio
   - Interval censoring

Recap

- Homework 4

2. Outline steps how you would show the following using simulation (no actual simulation is required)
   a) The Wilcoxon has the optimal power when the failure times are log-normally distributed, with equal variance in both groups but different means.
   b) With weights $\omega_i = S(t_i)$, the test is most powerful under the alternative hypothesis of log-logistic model

Recap

Group 1		Group 2		Power			
Mean	SD	Mean	SD	Logrank	Wilcoxon	Tarone-Ware	Peto-Prentice
1.5	1	2.5	1				
1.8	2	3.0	2				
1	1	3	2				
1	1	2	2				

- Homework 4 – Solution for Problem 2. a):

Step 1: Prepare a table shell or figure shell to compare the simulation results

Step 2: Choose parameters – parameters may need to be adjusted to obtain adequate power for comparisons

- Sample size n1 and n2 for Groups 1 and 2, respectively;
- Choose censoring distributions for Groups 1 and 2 – can use the same distribution
- Choose maximum follow-up time   Not follow up forever.
- Choose simulation trials, for power trials=2000 $2000 \times (n_1+n_2)$

Step 3: For each trial

- Generate data for two independent groups based on normal distributions, with sample size n1 and n2 for Groups 1 and 2, respectively
- Transform normally distributed data to log-normal
- Generate independent random censoring data using exponential distribution
- Obtain observed survival data
- Test between group difference in survival time and obtain the p-values for log-rank, Wilcoxon, Tarone-Ware, and Peto-Prentice

Step 4: Repeat Step 3 for 2000 times   2000 trials → 2000 test p-value (test statistic)

Step 5: Calculate the proportion of p-values that are significant at the level of 2-sided 0.05 (or 1-sided 0.025)

Recap

- Homework 4

3. Results from a trial evaluating a new treatment in comparison with a standard of care (SOC) indicate that biomark+ subgroup will have positive survival benefit with the new treatment and biomarker- subgroup actually gets harm.
    a) Please discuss what you think of an analysis by including all subject? Overall analyses including all subjects can be stratified or non-stratified analyses.
    b) How would you recommend the analysis?
    c) Do you think if FDA should approve this drug?

Recap

- Homework 4 - Solution for Problem 3

a) There are clear qualitative interaction between treatment groups and biomarker groups. The overall analyses will result in average treatment effects of the two biomarker groups.

  - The overall analysis without stratification is equivalent to use mixture distribution within each treatment group, which results in an averaged risk between the two biomarker strata. The averaged risk in each treatment group will depend upon the proportion of subjects distributed in each biomarker stratum in the group. The averaged risk will then be tested for the difference between treatments.
  - The overall analysis with stratification is to test the difference between the treatment groups first. The group differences for each stratum will then be averaged.
  - How similar the results of the two overall tests may depend upon the distribution of the biomarkers within each treatment group. If the distribution is balanced between the two treatment groups, the results of the two overall tests should be close to each other.

b) Pooled overall analyses should never be recommended, irrespective of stratified or non-stratified. The overall analysis is not helpful in interpreting the results. There should be separate analyses for each biomarker group.

c) Yes or No, Depends upon many factors

1. How strong is the evidence
2. If there is an unmet need
3. If there are safer drugs available
4. What is the risk/benefit ratio – the difference number needed to treat and number needed to harm
5. What is the prevalence of biomarker + and – in population
6. If there is a diagnostic companion test available to classify the biomarker status
7. What is the sensitivity and specificity of the diagnostic tests
8. Etc…

companion diagnostic test.

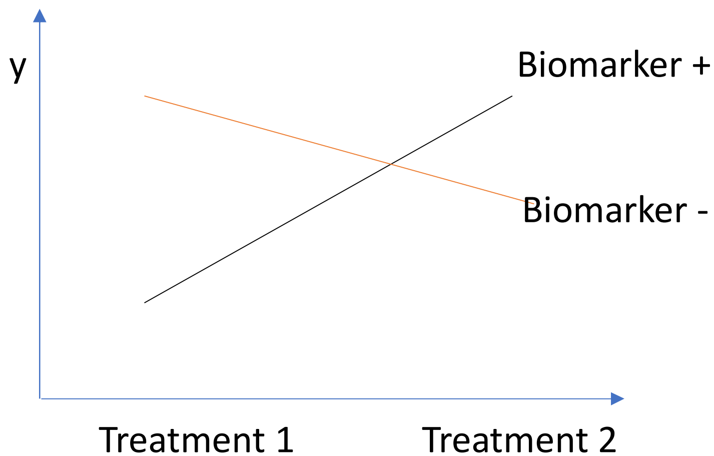
needed to treat $\frac{1}{10\%}$ have 1 patient to show benefit

needed to harm

Concept Review – Interaction Test

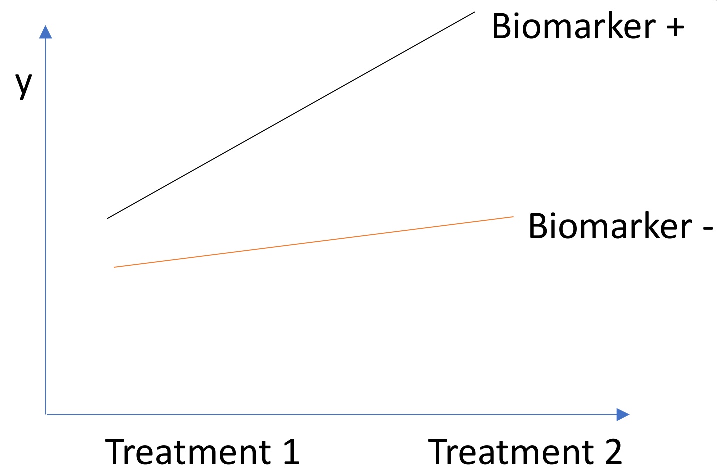
- Qualitative interaction

different direction



- Quantitative interaction

Same direction, different magnitude.



Test Interaction

- Let
  - λ_+ denote the hazard ratio for biomarker+ between Groups 1 and 2
  - λ_- denote the hazard ratio for biomarker- between Groups 1 and 2
- Hypothesis – should be two sided   interactions $\longrightarrow$ two-sided.

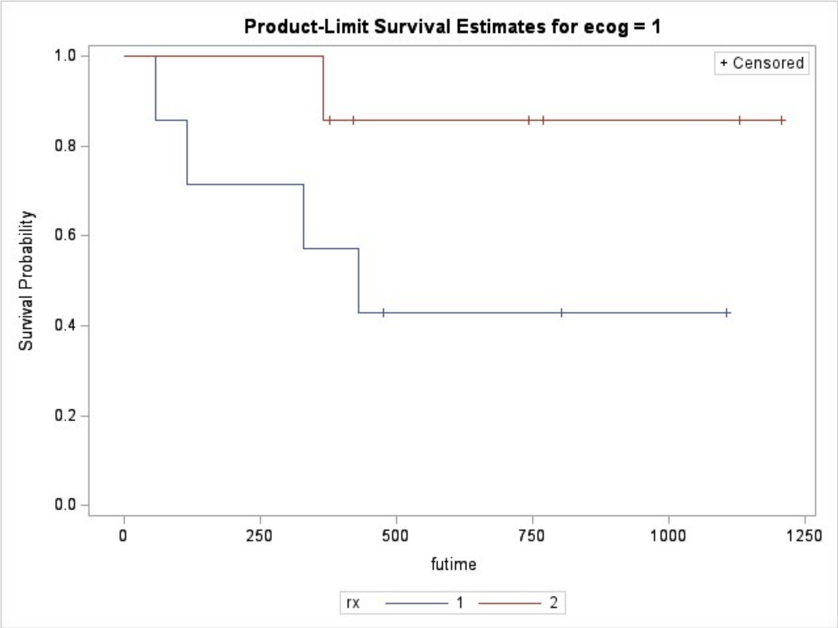
$$H_0: \lambda_+ = \lambda_- \ vs \quad H_A: \lambda_+ \neq \lambda_-$$

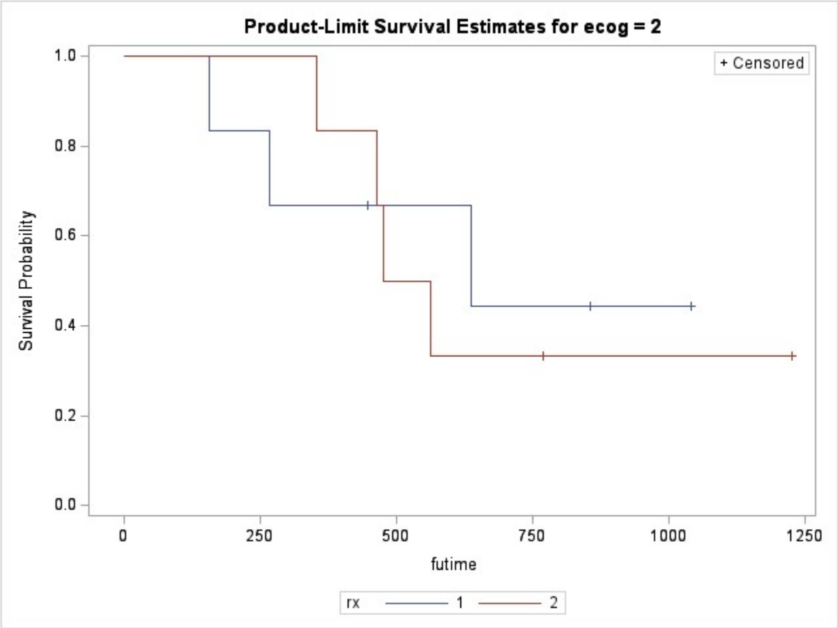
Example – Ovarian Data

	futime	fustat	age	resid.ds	rx	ecog.ps
1	59	1	72.3315	2	1	1
2	115	1	74.4932	2	1	1
3	156	1	66.4658	2	1	2
4	421	0	53.3644	2	2	1
5	431	1	50.3397	2	1	1
6	448	0	56.4301	1	1	2
7	464	1	56.9370	2	2	2
8	475	1	59.8548	2	2	2
9	477	0	64.1753	2	1	1
10	563	1	55.1781	1	2	2
11	638	1	56.7562	1	1	2
12	744	0	50.1096	1	2	1
13	769	0	59.6301	2	2	2
14	770	0	57.0521	2	2	1
15	803	0	39.2712	1	1	1
16	855	0	43.1233	1	1	2
17	1040	0	38.8932	2	1	2
18	1106	0	44.6000	1	1	1
19	1129	0	53.9068	1	2	1
20	1206	0	44.2055	2	2	1
21	1227	0	59.5890	1	2	2
22	268	1	74.5041	2	1	2
23	329	1	43.1370	2	1	1
24	353	1	63.2192	1	2	2
25	365	1	64.4247	2	2	1
26	377	0	58.3096	1	2	1

- Dataset available in R survival package
  - futime: survival or censoring time (day)
  - fustat: censoring status (censor=0)
  - age: in years
  - resid.ds: residual disease present (1=no,2=yes)
  - rx: treatment group
  - ecog.ps: ECOG performance status (1 is better, see reference)
- Ecog performance status as strata

Example- Ovarian Data





Example Ovarian Data

ecog. = 1

= 2

- Rank statistics log-rank

ECOG	Rank Statistics	Variance	P-value
1	1.758	1.222	0.112
2	-0.258	1.707	0.843

- Interaction test

$$\frac{1.758+0.258}{\cancel{\sqrt{1.222}+\sqrt{1.707}}} = 0.836$$

$$\sqrt{1.222+1.707}$$

p-value $\left\{ \begin{array}{l} 0.05 \\ 0.10 \\ 0.15 \end{array} \right.$

If the test statistics is normally distributed, 2-sided p-value=0.403

Multiplicity Discussion

- Sources of multiplicity
- How to handle in clinical trials

multiple test

Syllabus

1. Introduction
   - Survival data
   - Censoring mechanism
   - Application in medical field
2. Concepts and definitions
   - Survival function
   - Hazard function
3. Non-parametric approach
   - Life table
   - Kaplan-Meier survival estimate
   - Hazard function
   - Median and percentile survival time
4. Hypothesis testing
   - Overview – hypothesis, test statistics, p-values
   - Log-rank
   - Wilcoxon
   - Gehan test
5. **Study design and sample size estimation**
   - **Overview**
   - **Survival sample size estimation**
   - **Accrual time and Study duration**
6. Semiparametric model – proportional hazard model
   - Partial likelihood
   - Inference
   - Time varying covariates
   - Stratification
7. Model checking in the PH model
   - Model checking
   - Residuals
8. Parametric model
   - Parametric proportional hazard model
   - Accelerate failure model
9. Other topics
   - Competing risk
   - Recurrent events
   - Non-proportional hazard ratio
   - Interval censoring

Outline

- One sample
- Two samples
- Software for survival sample size calculation
  - R package Library(gsdesign)
  - East
  - nQuery
  - PASS
  - SAS

Why Do We Need Sample Size Calculation

- Determine the success rate of a trial
  - Based on the expected treatment effect
- Estimate the duration of the trial
- Determine the cost of the trial
- Planning the trial
  - How many sites are needed?

- The very first step in planning a trial…

Review of Sample Size Calculation

from previous literature/trial
ALL ASSUMPTIONS

- Hypothesis testing
  - Two sample comparison
    - Group 1 (Treatment) has n_1 subjects, $x_1 \sim N(\mu_1, \sigma^2)$
    - Group 2 (Control) has n_2 subjects, $x_2 \sim N(\mu_2, \sigma^2)$
    - $\delta = \mu_1 - \mu_2$

$\mu_1 - \mu_2 = \delta \rightarrow$ difference δ, diff n

  - Two-sided hypotheses

Two-sided makes no sense $\mu_1 > \mu_2$ (benifitial)

    - H_0: $\mu_1 = \mu_2$ vs H_A: $\mu_1 \neq \mu_2$
  - One-sided hypotheses
    - H_0: $\mu_1 \leq \mu_2$ vs H_A: $\mu_1 > \mu_2$

Elements Needed in Sample Size Calculation

- Two error rates
  - α • Type I – significance level $Pr\{\text{reject } H_0 \mid H_0 \text{ is true}\}$
  - β • Type II – determines power $Pr\{\text{fail to reject } H_0 \mid H_a \text{ is true}\}$
- Assumptions
  - Treatment effect in each treatment groups
  - Variance
  - Dropout rate
- More needed in survival analysis

Review of Sample Size Calculation

- Control of error rates
  - Type I error α – reject null when null is true
  - Type II error β – fail to reject null when alternative is true

- In drug development and evaluation
  - α – claim an ineffective drug to be efficacious, controlled at the 1-sided level of 0.025 in individual trial
  - β – fail to bring an efficacious drug to patients
  - *1-* β – study power

Review of Sample Size Calculation

- Specify the significance level - the type I error

$\alpha = 0.05$

- For Phase III studies
    - α=0.05 for two-sided tests
    - α=0.025 for 1-sided tests → chi-sq
  - Phase II studies – proof of concept studies
    - α=0.05 to 0.15 for two-sided tests

Q: Why could be higher in phase-II.

A: Phase-III → comfermative → stringent.

- Determine critical value for 1-sided hypothesis
  - Need to understand the distribution of the test statistics U
    - Often normally distributed
  - Determine critical value $c_{1-\alpha},\ P(U \geq c_{1-\alpha}) = \alpha$
  - If $U \sim N(0,1), c_{1-\alpha} = z_{1-\alpha}$

$\alpha = 0.05$

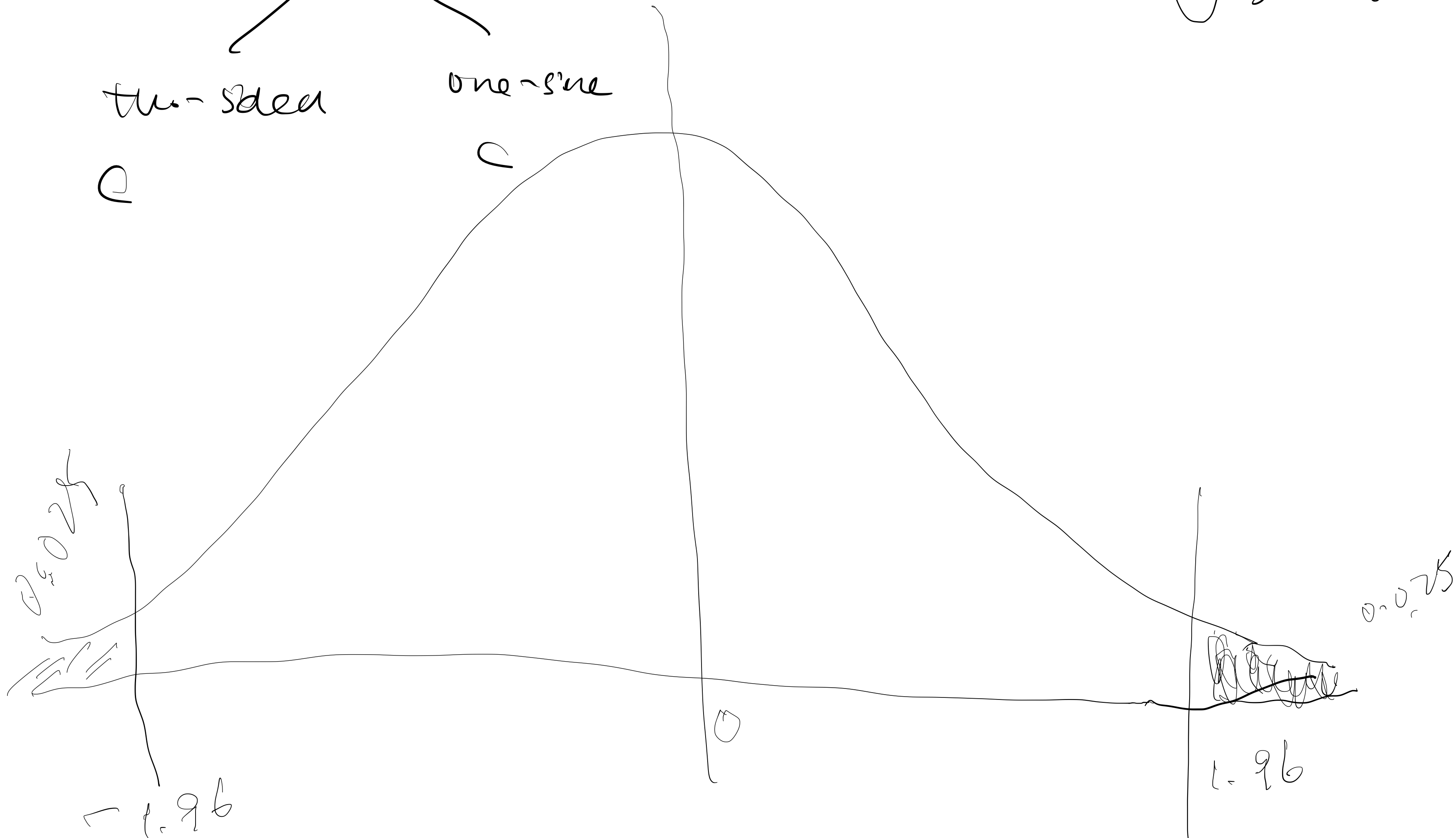
two-sided

c

one-side

c

$U > C_{1-\alpha}$

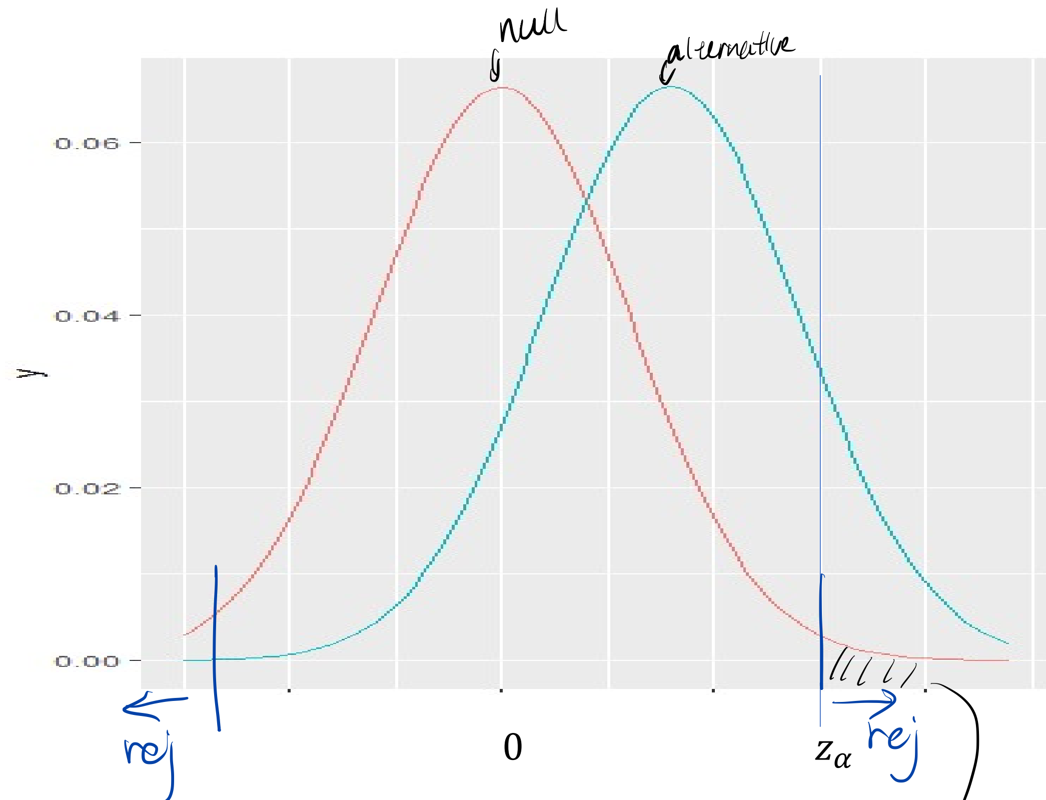


1. ~~Dese~~ Decide two/one side.
2. α.

Review of Sample Size Calculation

- How to choose power
  - Depending upon the expected treatment effect – assumption
  - Depending upon resource
  - $\geq 85\%$ in Phase III
  - About 80% or even lower in early phases

Review of Sample Size Calculation



Relationship between power and significance level

Red curve : Distribution of Z-statistic $\sim N(0,1)$

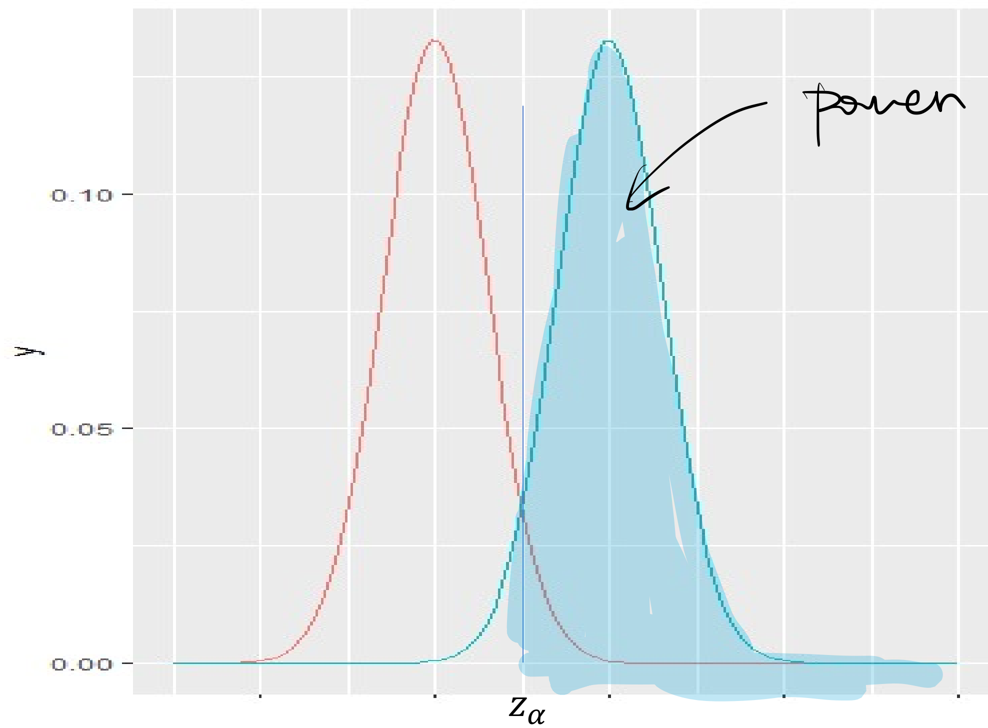
critical value: $1.96 \rightarrow \alpha = 0.025$

$1.28 \rightarrow \alpha = 0.010$

$P(\text{tail red}) = \alpha$.

$P(\text{tail green}) = 1 - \beta = \text{power}$

Review of Sample Size Calculation



Review of Sample Size

- Let $\bar{x}_i = \frac{\sum_{j=1}^{n_i} x_{ij}}{n_i}$, $s^2 = \frac{\sum_{i=1}^{2}\sum_{j=1}^{n_i}(x_{ij}-\bar{x})}{n_1+n_2-1}$ $i = 1,2$
- $U = \frac{\bar{x}_1-\bar{x}_2}{s\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}$,
- Let $n_1 = n_2$ for equal randomization ratio
- Under null $U = \frac{(\bar{x}_1-\bar{x}_2)\sqrt{n}}{s\sqrt{2}} \sim N(0,1)$
- Select critical value $z_{1-\alpha}$, $P_{H_0}(U \geq z_{1-\alpha}) = \alpha$

Review of Sample Size

assumption, e.g. 30% percent change.

- Under alternative $U \sim N(\frac{\delta\sqrt{\frac{n}{2}}}{\sigma}, 1)$ standardized.

$$1-\beta = P_{H_A}(U \geq z_{1-\alpha}) = P_{H_A}\left(U - \frac{\delta\sqrt{\frac{n}{2}}}{\sigma} \geq z_{1-\alpha} - \frac{\delta\sqrt{\frac{n}{2}}}{\sigma}\right)$$

$$= P_{H_A}\left(U - \frac{\delta\sqrt{\frac{n}{2}}}{\sigma} \geq z_{1-\alpha} - \frac{\delta\sqrt{\frac{n}{2}}}{\sigma}\right) = 1 - \phi(z_{1-\alpha} - \frac{\delta\sqrt{\frac{n}{2}}}{\sigma})$$

$$z_\beta = z_{1-\alpha} - \frac{\delta\sqrt{\frac{n}{2}}}{\sigma}$$

$$n = \frac{2(z_{1-\alpha} - z_\beta)^2}{\delta^2}\sigma^2 = \frac{2(z_{1-\alpha} + z_{1-\beta})^2}{\delta^2}\sigma^2$$

$$(z_{1-\alpha} - z_\beta) = \frac{\delta\sqrt{\frac{n}{2}}}{\sigma}$$

$$\frac{\sigma(z_{1-\alpha} - z_\beta)}{\delta} = \sqrt{\frac{n}{2}}$$

$$\frac{2\sigma^2(z_{1-\alpha} - z_\beta)}{\delta^2}$$

Sample Size for Survival Analyses

- The hypotheses:

$$H_0\colon S_1(t) \leq S_0(t) \qquad \text{vs} \qquad H_A\colon S_1(t) > S_0(t)$$

- In addition to the required significance level and power, the sample size is dependent of
  - Number of events needed
  - Event rates
  - Follow-up time
  - Accrual rate
  - Censoring distribution

- Survival data distribution is assumed
  - Survival time distributions
    - T $\sim \exp(\lambda)$
    - Piece-wise exponential distribution

Sample Size for Survival Analyses

- Survival time is usually quantified by
  - Median survival time
  - Hazard rates

> if we know the family of distribution, both are equivalent.

- Assuming the survival time follows exponential distribution
  - Constant hazard rate
  - $S_0(t) = e^{-\lambda_0 t}$ $\quad S_1(t) = e^{-\lambda_1 t}$
  - H_0: $\lambda_0 \le \lambda_1$ vs H_A: $\lambda_0 > \lambda_1$

$\Leftrightarrow$ $H_0: S_0(t) \le S_1(t)$ $\qquad$ $H_A: S_0(t) > S_1(t)$

One Sample Survival Analysis

- Single arm studies are often seen in oncology clinical trials
  - Late lines of therapies already very sick, no need for control.
  - Phase II studies
- Often the endpoints are
  - Overall response rates Binary.
  - Complete response rates
  - PFS
  - Overall survival

One Sample Survival Analysis

- No Censoring
- d – number of events, all subjects have events
- $T_i \sim Exp(h), i = 1, \ldots, d$
  - $\bar{T} \sim N\left(\frac{1}{h}, \frac{1}{dh^2}\right)$ CLT.

why we need this transformation $\hat{h} = \frac{1}{\bar{T}}$ estimated. hazard rate.

- Taking variance-stabilizing transformation, by delta method,
  - Distr. of $\hat{h}$ $\log \hat{h} = -\log \bar{T} \sim N(\log h, \frac{1}{d})$ Variance has nothing to do with the variance.
  - Variance is the inverse of number of events

One Sample Survival Analysis

- $H_0: h \geq h_0$ versus $H_a: h < h_0$

h_0 may come from historical trials.
"standard care" hazard.

- $\sqrt{d}\ \left(\log \hat{h} - \log h_0\right) \sim N(0,1)$

- Total number of events needed to obtain power $1 - \beta$ at the significance level of α

$$\underbrace{\text{sample size}}_{\text{all sub. have events}} = d = \frac{(z_{1-\alpha} - z_\beta)^2}{(\log \lambda)^2} \text{ where } \lambda = \frac{h}{h_0}$$

$(\log h - \log h_0)^2$

Example – One Sample

- To achieve 80% power at the 1-sided significant level of 0.025 to detect a 50% increase in median PFS from 12 months in historical control to 18 months

- Assuming exponential distribution
- $\lambda = \frac{12}{18} = \frac{2}{3}$ convertion from hazard ratio to ratio of median survival time
- $d = \frac{(z_{1-\alpha} - z_\beta)^2}{(\log \lambda)^2} = \frac{(1.96+0.84)^2}{(\log 0.67)^2} = 48$ better to ceilling the value.

One Sample Survival Analysis

- Number of subjects
- Assuming
  - Everyone follow-up time τ
  - Event rate is $Pr = 1 - S(\tau)$ expected event rate at the end of follow-up.
  - The number of subjects N needed will be $N = d/(1 - S(\tau))$

Example – One Sample

- To achieve 80% power at the 1-sided significant level of 0.025 to detect a 50% increase in median PFS from 12 months in historical control to 18 months
  - Follow-up period is 36 month
  - Assuming exponential distribution
  - $\lambda = \frac{12}{18} = \frac{2}{3}$
  - $d = \frac{(z_{1-\alpha} - z_{\beta})^2}{(\log \lambda)^2} = \frac{(1.96+0.84)^2}{(\log 0.67)^2} = 48$
  - Hazard function $h = \frac{-\log 0.5}{18} = 0.0385$
  - Event rate $Pr = 1 - S(36) = 1 - e^{-0.0385*36} = 0.75$
  - Subjects enrolled $N = \frac{48}{0.75} = 64$

$\phi \quad e^{-h \cdot t} = 0.5$

$e^{-h \cdot 18} = 0.5$

$S_t = 0.5$

$-h\,18 = \log 0.5$

$h = \frac{-\log 0.5}{18}$

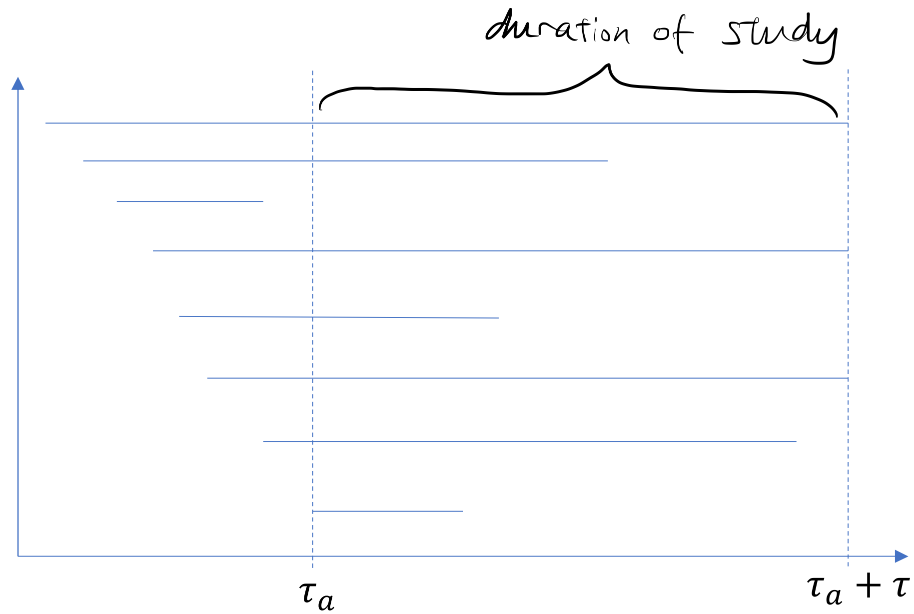
One Sample Survival Analysis

- Assuming
  - A uniform accrual period τ_a - staggered enrollment
  - Minimum follow-up time τ
  - Expected event rate is

$$Pr = \int_0^{\tau_a} \left(1 - S(\tau + a)\right) f(a) da$$

$$= 1 - \frac{\int_0^{\tau_a} S(\tau + a) da}{\tau_a}$$

$$= 1 - e^{-h\tau}(1 - e^{-h\tau_a})/h\tau_a$$



Example – One Sample

- To achieve 80% power at the 1-sided significant level of 0.025 to detect a 50% increase in median PFS from 12 months in historical control to 18 months
  - 12-month accrual period uniform distribution
  - Follow-up period is 36 month
  - Assuming exponential distribution
  - $\lambda = \frac{12}{18} = \frac{2}{3}$
  - $d = \frac{(z_{1-\alpha} - z_{\beta})^2}{(\log \lambda)^2} = \frac{(1.96+0.84)^2}{(\log 0.67)^2} = 48$
  - Hazard function $h = \frac{-\log 0.5}{18} = 0.0385$
  - Expected event rate $Pr = 1 - e^{-h\tau}(1 - e^{-h\tau_a})/h\tau_a = 0.8$
  - Subjects enrolled $N = \frac{48}{0.8} = 60$

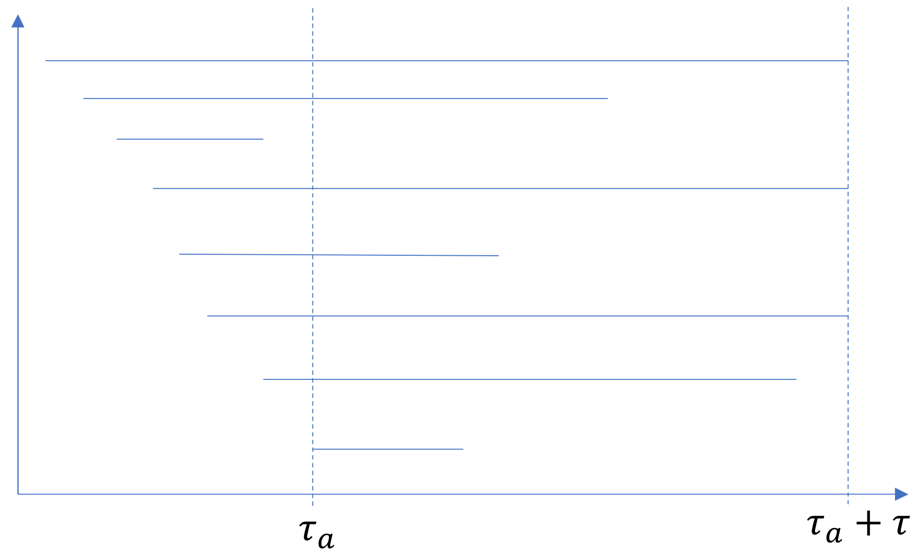
One Sample Survival Analysis

- Assuming
  - A uniform accrual period τ_a - staggered enrollment
  - Minimum follow-up time τ
  - Independent censoring follows $C \sim \exp(h_c)$
  - Expected event rate is

$$Pr = \frac{h}{h + h_c} \int_0^{\tau_a} \big(1 - S(\tau + a)\big) f(a) da$$

$$= (1 - \frac{\int_0^{\tau_a} S(\tau + a) da}{\tau_a}) \frac{h}{h + h_c}$$

$$= (1 - \frac{e^{-(h+h_c)\tau}(1 - e^{-(h+h_c)\tau_a})}{(h+h_c)\tau_a}) \frac{h}{h+h_c}$$



Example – One Sample

- To achieve 80% power at the 1-sided significant level of 0.025 to detect a 50% increase in median PFS from 12 months in historical control to 18 months
  - 12-month accrual period uniform distribution
  - Follow-up period is 36 month
  - Assuming $C \sim \exp(0.002)$
  - $\lambda = \frac{12}{18} = \frac{2}{3}$
  - $d = \frac{(z_{1-\alpha} - z_{\beta})^2}{(\log \lambda)^2} = \frac{(1.96+0.84)^2}{(\log 0.67)^2} = 48$
  - Hazard function $h = \frac{-\log 0.5}{18} = 0.0385, h_c = 0.002$
  - Expected event rate $Pr = (1 - \frac{e^{-(h+h_c)\tau}(1-e^{-(h+h_c)\tau_a})}{(h+h_c)\tau_a}) \frac{h}{h+h_c}$ =0.775
  - Subjects enrolled $N = \frac{48}{0.775} = 62$

Two Samples Survival Analysis

- Testing two survival functions
  - Control $T_0 \sim \exp(\, h_0)$
  - Treatment $T_1 \sim \exp(\, h_1)$,
- Hypotheses, $H_0\colon\ h_0 \le h_1$ vs $H_A\colon\ h_0 > h_1$
- Hazard ratio $\lambda = h_1 /\, h_0$
- By the central limit theorem, variance stabilizing transformation, and the delta method

$$\log \bar{T}_0 - \log \bar{T}_1 \sim N(\log \lambda\,, \frac{1}{d_1} + \frac{1}{d_2})$$

$$\frac{\log(\bar{T}_0/\bar{T}_1)}{\sqrt{\left(\frac{1}{d_1} + \frac{1}{d_2}\right)}} \sim N(0{,}1) + \log \lambda \,/ \sqrt{(\frac{1}{d_1} + \frac{1}{d_2})}$$

Two Samples Survival Analysis

1-sided, $\alpha = 0.025$
2-sided $\alpha = 0.05$.

Set critical value $z_{1-\alpha}$ and achieving the desired power

$$1-\beta = 1-\phi(z_\beta)$$

$$= P_{H_A}\left(\underbrace{log(T_0/T_1)/\sqrt{\left(\frac{1}{d_1}+\frac{1}{d_2}\right)} - \log\lambda\,/\sqrt{\left(\frac{1}{d_1}+\frac{1}{d_2}\right)}}_{\text{Under null} \to 0} \geq z_{1-\alpha} - \log\lambda\,/\sqrt{(\frac{1}{d_1}+\frac{1}{d_2})} \right)$$

$$= 1-\phi\left(z_{1-\alpha} - \frac{\log\lambda}{\sqrt{\left(\frac{1}{d_1}+\frac{1}{d_2}\right)}} \right)$$

$$z_\beta = z_{1-\alpha} - \log\lambda\,/\sqrt{(\frac{1}{d_1}+\frac{1}{d_2})}$$

Two Samples Survival Analysis

$$z_\beta = z_{1-\alpha} - \log\lambda \,/\, \sqrt{(\frac{1}{d_1} + \frac{1}{d_2})} = z_{1-\alpha} - \log\lambda \Big/ \sqrt{\frac{4}{d}}$$

technique

Assumption

One simple scenario is to let $d_1 = d_2 = \frac{d}{2}$, total number of events are

$$d = 4(\frac{z_{1-\alpha} - z_\beta}{\log\lambda})^2$$

May calculate each group's event

- if there is a treatment difference, generally, $d_1 \neq d_2$

Two Samples Survival Analysis

- Including the trial design considerations
  - A uniform accrual period τ_a - staggered enrollment → uniform distribution $(0, \tau_a)$
  - Minimum follow-up time τ Total follow up time $\tau + \tau_a$
  - Independent censoring follows
    - $C_0 \sim \exp(h_{C_0})$ $C_1 \sim \exp(h_{C_1})$
  - Expected event rate is for group $i, i = 0,1$ h_0

Pr_1

Pr_0

$$Pr_i = \frac{h_i}{h_i + h_{C_i}} \int_0^{\tau_a} \left(1 - S(\tau + a)\right) f(a) da$$

$$= \left(1 - \frac{\int_0^{\tau_a} S(\tau + a) da}{\tau_a}\right) \frac{h_i}{h_i + h_{C_i}}$$

$$= \left(1 - \frac{e^{-(h_i + h_{C_i})\tau}\left(1 - e^{-(h_i + h_{C_i})\tau_a}\right)}{(h_i + h_{C_i})\tau_a}\right) \frac{h_i}{h_i + h_{C_i}}$$

$n_1 = \dfrac{d_1}{Pr_1} = \dfrac{N}{2} = \dfrac{d_2}{Pr_2} = n_2$

$d_1 + d_2 = d$

$\tau_a \uparrow$

sample size $\dfrac{d_i}{Pr_i}$

Example – Two Sample Survival

- A drug company is developing a new treatment for multiple myeloma. A randomized and controlled study is designed with randomization ratio 1:1
  - 90% power
  - 1-sided significance level of 0.025
  - Endpoint is progression-free-survival
- Assuming the median survival time for the standard of care is 9 months, and the median survival time for the new treatment is 12 months
  - Hazard rate $\frac{-\log(0.5)}{9}$
    - $h_0 = \frac{\log 2}{9} = 0.077$
    - $h_1 = \frac{\log 2}{12} = 0.058$
  - Hazard ratio λ is 9/12=0.75, logλ =-0.286
  - Expecting 25% risk reduction
- The total number of events needed is
  - $d = 4(\frac{z_{1-\alpha} - z_\beta}{\log 0.75})^2 = 4(\frac{1.96+1.28}{-0.286})^2 \approx 509$

Scenario	Group	d	tao	tao-a	hi	hci	pri	N	N/2
1	0	508 509	24	18	0.077	0.002	0.90	568	**284**
	1	509	24	18	0.058	0.002	0.83	616	**308**
2	0	507 509	36	18	0.077	0.002	0.94	539	**269**
	1	509	36	18	0.058	0.002	0.90	567	**283**
3	0	509	24	12	0.077	0.002	0.88	578	**289**
	1	509	24	12	0.058	0.002	0.80	634	**317**

587.51

551.74 264

Example – Two Sample Survival

N=587.51

- Scenario I
  - A uniform accrual period $\tau_a = 18$ months
  - Minimum follow-up time $\tau = 24$ months
  - Independent censoring follows
    - $C_0 \sim \exp(0.002)$ and $C_1 \sim \exp(0.002)$

  Number of subjects per group N=272

- Scenario II
  - A uniform accrual period $\tau_a = 18$ months
  - Minimum follow-up time $\tau =$ ~~24~~ 36 months
  - Independent censoring follows
    - $C_0 \sim \exp(0.002)$ and $C_1 \sim \exp(0.002)$

  Number of subjects per group N=282

- Scenario III
  - A uniform accrual period $\tau_a = 12$ months
  - Minimum follow-up time $\tau = 24$ months
  - Independent censoring follows
    - $C_0 \sim \exp(0.002)$ and $C_1 \sim \exp(0.002)$

  Number of subjects per group N=286

$$\frac{\frac{N\,\cancel{N^2}}{2\,\cancel{4}} P_{r1}P_{r2}}{\frac{\cancel{N}}{\cancel{8}}(P_{r1}+P_{r2})} =$$

$$\frac{d_1 d_2}{d_1 + d_2} = \left(\frac{(1.96+1.28)}{\log(0.75)}\right)^2$$

$$\frac{d_1}{P_{r1}} = \frac{d_2}{P_{r2}} = \frac{N}{2}$$

$$d_1 = \frac{N}{2} P_{r1} \quad d_2 = \frac{N}{2} P_{r2}$$

−1.28 1.96 0.75

$$z_\beta = z_{1-\alpha} - \log\lambda \Big/ \sqrt{\left(\frac{1}{d_1} + \frac{1}{d_2}\right)}$$

The SAS System

The POWER Procedure
Log-Rank Test for Two Survival Curves

Fixed Scenario Elements	
Method	Lakatos normal approximation
Accrual Time	18
Follow-up Time	36
Reference Survival Curve	SOC
Form of Survival Curve 1	Exponential
Form of Survival Curve 2	Exponential
Hazard Ratio	1.33
Group 1 Loss Exponential Hazard	0.002
Group 2 Loss Exponential Hazard	0.002
Nominal Power	0.9
Number of Sides	2
Number of Time Sub-Intervals	12
Alpha	0.05
Group 1 Weight	1
Group 2 Weight	1

Computed Ceiling Event Total		
Fractional Event Total	Actual Power	Ceiling Event Total
531.217786	0.900	532

SAS Code – Comparing Two Survival Curves

```
proc power;
   twosamplesurvival test=logrank
      curve("SOC")=9:0.5
         refsurvival="SOC"
      hazardratio=1.5
      accrualtime = 2
      followuptime = 3
grouplossexphazards=0.002|0.002
      power = 0.8
         eventstotal=.
         /*ntotal=. */
      /*npergroup = .*/;
run;
```

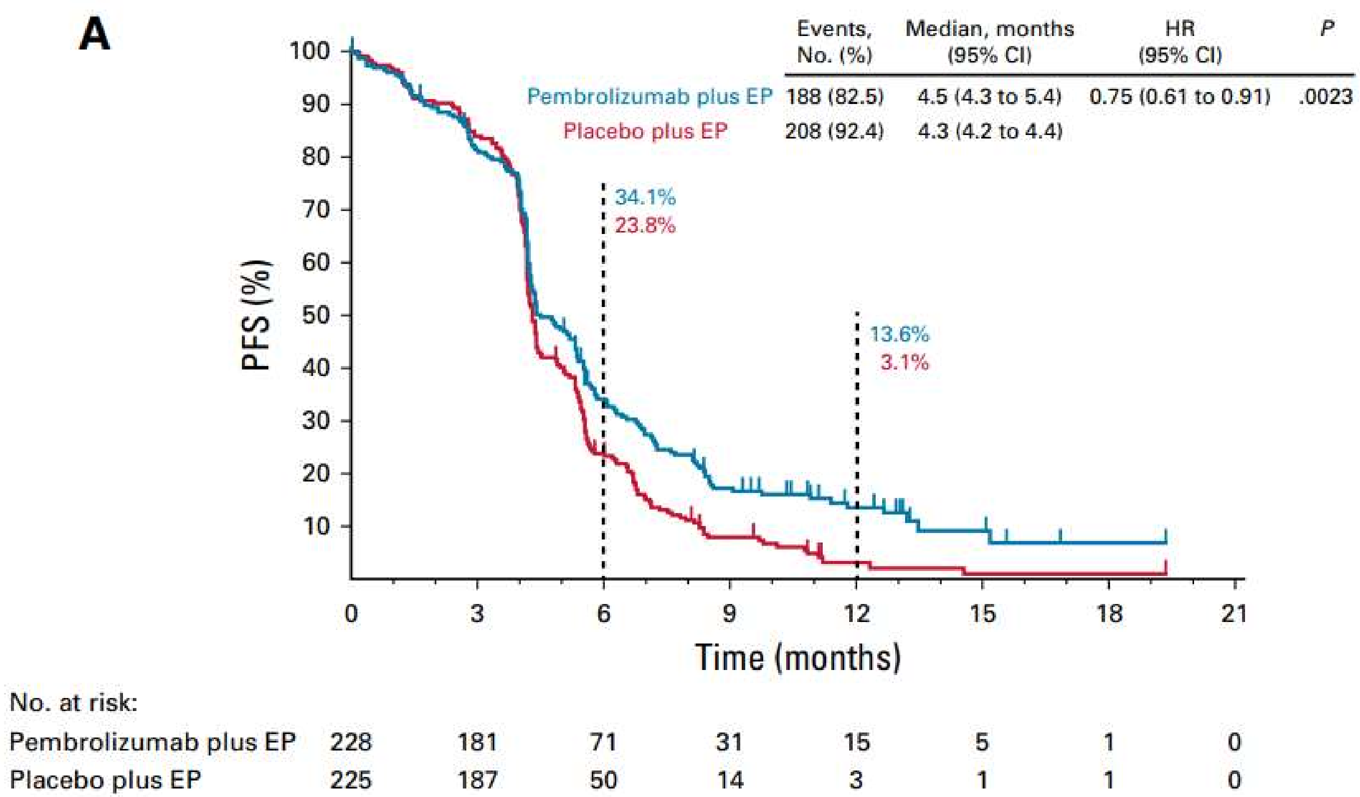
The POWER Procedure
Log-Rank Test for Two Survival Curves

Fixed Scenario Elements	
Method	Lakatos normal approximation
Accrual Time	2
Follow-up Time	3
Reference Survival Curve	SOC
Form of Survival Curve 1	Exponential
Form of Survival Curve 2	Exponential
Hazard Ratio	1.5
Group 1 Loss Exponential Hazard	0.002
Group 2 Loss Exponential Hazard	0.002
Nominal Power	0.8
Number of Sides	2
Number of Time Sub-Intervals	12
Alpha	0.05
Group 1 Weight	1
Group 2 Weight	1

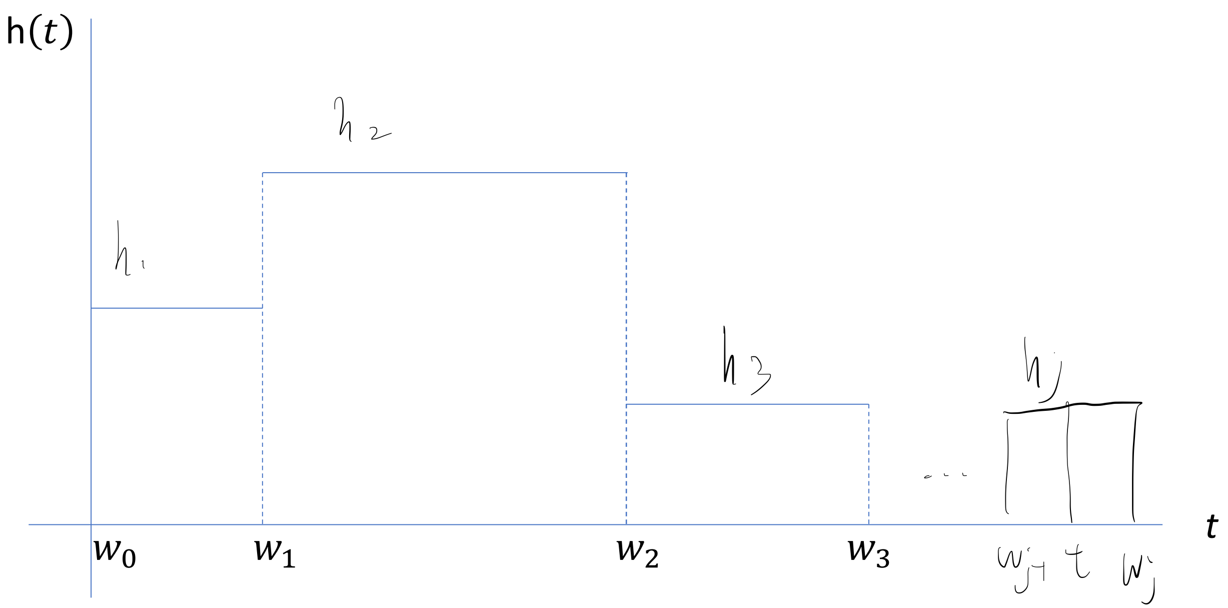
Computed Ceiling Event Total		
Fractional Event Total	Actual Power	Ceiling Event Total
190.680297	0.801	191

Simulation Based Sample Size Calculation

- So far, we have assumed exponential distribution
  - Constant hazard rates
  - Constant hazard ratio
- Often, we see clear violations of the assumptions
- Sample size might be under powered based on exponential assumption
- Simulation can be
  - based on piece-wise exponential
  - Accrual rate can change as well



Recall: Piecewise Hazard Function



Piecewise Exponential –The Cumulative Hazard Function

- The hazard function

$$\mathrm{h}(t) = h_1 I(t \le w_1) + h_2 I(w_1 < t \le w_2) + \cdots$$

where $w_1, w_2, \ldots,$ are fixed time intervals, $w_0 = 0$

- At time $t \in (w_{j-1}, w_j)$, the cumulative hazard function can be written as

$$H(t) = \sum_{i<j} h_i (w_i - w_{i-1}) + h_j \left(t - w_{j-1}\right) \mathrm{I}(t \in (w_{j-1}, w_j))$$

Need to remember

Piecewise Exponential – The Survival Function

- Recall $S(t) = e^{-H(t)}$
- Therefore, for $t \in (\overset{w_{j-1}}{\cancel{w_j}}, \overset{w_j}{\cancel{w_{j+1}}})$

$$S(t) = e^{-\{\sum_{i<j} h_i(w_i - w_{i-1}) + h_j(t - w_{j-1})\}}$$

$$= \prod_{i<j} e^{-h_i(w_i - w_{i-1})}\, e^{-h_j(t - w_{j-1})}$$

Piecewise Exponential - PDF

- Recall $f(t) = -\frac{dS(t)}{dt}$
- Therefore, for $t \in (\overset{w_{j-1},\, w_j}{\cancel{w_j, w_{j+1}}})$

$$f(t) = -\frac{d}{dt} e^{-\{\sum_{i<j} h_i(w_i - w_{i-1}) + h_j(t - w_{j-1})\}}$$

$$= h_j \prod_{i<j} e^{-h_i(w_i - w_{i-1})}\, e^{-h_j(t - w_{j-1})}$$

Simulation Based Sample Size Calculation

- Another reason to use simulation
  - No closed formulations for weighted log-rank tests

Test	Weight ω_i
Log-rank	$\omega_i = 1$
Gehan's Wilcoxon	$\omega_i = n_i$
Peto/Prentice	$\omega_i = S(t_i)$
Fleming-Harrington	$\omega_i = S(t_{i-1})^{\rho}(1 - S(t_{i-1}))^{q} \;\; \rho, q \geq 0$
Tarone-Ware	$\omega_i = \sqrt{n_i}$

Simulation Procedures for Power

- Specify
  - Set the significance level
  - Randomization ratio and number of subjects in each treatment group
  - Hazard functions for each treatment groups
  - Dropout hazard functions for each treatment groups
  - Generate accrual patterns
- Iteration steps to generate N trials, in each trial
  - Generate start time for each subject based on the accrual patterns
  - Allocate subjects in each treatment group based on the randomization scheme
  - Generate event time for subjects in each treatment group
  - Generate censor time for each subject
  - Obtain the observed survival time
  - Apply the tests
- The proportion that rejects the null is the power

Simulation Based Sample Size Calculation

```
proc power;
  twosamplesurvival test=logrank
    curve("Standard") = 5 : 0.5
    curve("Proposed") =
      (1 to 5 by 1):(0.95 0.9 0.75 0.7 0.6)
    groupsurvival = "Standard" | "Proposed"
    accrualtime = 2
    followuptime = 3
    groupmedlosstimes = 10 | 20 5
      /* median loss time*/
    power = 0.8
    npergroup = .;
run;
```

The SAS System

The POWER Procedure
Log-Rank Test for Two Survival Curves

Fixed Scenario Elements	
Method	Lakatos normal approximation
Accrual Time	2
Follow-up Time	3
Group 1 Survival Curve	Standard
Form of Survival Curve 1	Exponential
Group 2 Survival Curve	Proposed
Form of Survival Curve 2	Piecewise Linear
Group 1 Median Loss Time	10
Nominal Power	0.8
Number of Sides	2
Number of Time Sub-Intervals	12
Alpha	0.05

Computed N per Group			
Index	Median Loss Time 2	Actual Power	N per Group
1	20	0.800	228
2	5	0.801	234

Simulation Based Sample Size Calculation

```
proc power;
  twosamplesurvival test=logrank
    curve("Standard") = 5 : 0.5
    curve("Proposed") =
      (1 to 5 by 1):(0.95 0.9 0.75 0.7 0.6)
    groupsurvival = "Standard" | "Proposed"
    accrualtime = 2
    followuptime = 3
    groupmedlosstimes = 10 | 20 5
     /* median loss time*/
    power = 0.8
    eventstotal = .;
run;
```

Computed Ceiling Event Total				
Index	Median Loss Time 2	Fractional Event Total	Actual Power	Ceiling Event Total
1	20	167.680114	0.801	168
2	5	171.849127	0.800	172

Compare Sample Sizes

- Comparison of sample sizes
- Hazard ratio $\lambda = 1.5$ Constant rate, log-rank reaches its optimal power
- Power=0.8
- 1-side significance level of 0.025
- 1:1 randomization ratio
- Explain the difference of the sample size

less N

	Log-rank	Gehan	Tarone-ware
Events	191	202	195
Subjects per group	302	319	308

Example – Piecewise Exponential

- A drug company is developing a new treatment for disease X. A randomized and controlled study is designed with randomization ratio 1:1 Always give the largest power when N_{total} is fixed.
  - 90% power
  - 1-sided significance level of 0.025
  - Endpoint is progression-free-survival
- Assuming
  - No treatment difference in the first 6 months
  - The median survival time for the first 6 months
    - Standard of care is 8 months
    - New treatment is 8 months
  - The median survival time after 6 months
    - Standard of care is 9 months
    - New treatment is 12 months
- Planned accrual period is 12 months
- Minimum follow-up is 24 months

- very agressive trial

Example – Piecewise Exponential

$1 - \text{pexp}\left(6, \frac{\log(2)}{8}\right)$

```
proc power;
   twosamplesurvival test=logrank
      curve("SOC") = 6:0.594 36:0.059
      curve("New Treatment") = 6:0.594
36:0.187
      groupsurvival = "SOC" | "New
Treatment"
      accrualtime = 12
      followuptime = 24
      groupmedlosstimes = 30 20 | 30 20  /*
median loss time*/
      power = 0.9
          eventstotal=.
          /*ntotal=.*/
      /*npergroup = .*/;
run;
```

The POWER Procedure
Log-Rank Test for Two Survival Curves

Fixed Scenario Elements	
Method	Lakatos normal approximation
Accrual Time	12
Follow-up Time	24
Group 1 Survival Curve	SOC
Form of Survival Curve 1	Piecewise Linear
Group 2 Survival Curve	New Treatment
Form of Survival Curve 2	Piecewise Linear
Nominal Power	0.9
Number of Sides	2
Number of Time Sub-Intervals	12
Alpha	0.05
Group 1 Weight	1
Group 2 Weight	1

Computed Ceiling Event Total					
Index	Median Loss Time 1	Median Loss Time 2	Fractional Event Total	Actual Power	Ceiling Event Total
1	30	30	1855.436477	0.900	1856
2	30	20	2200.321524	0.900	2201
3	20	30	2275.449077	0.900	2276
4	20	20	2632.218210	0.900	2633

$$=\left(1-\frac{e^{-(h_i+h_{c_i})\tau}\left(1-e^{-(h_i+h_{c_i})\tau_a}\right)}{(h_i+h_{c_i})\tau_a}\right)\frac{h_i}{h_i+h_{c_i}}$$

$$h_0 = -\frac{\log 0.5}{20} \qquad h_1 = -\frac{\log 0.5}{28} \qquad \lambda = \frac{20}{28} = \frac{5}{7}$$

$C = 0.002$

$\tau = 24$

$\tau_a = 18$

$$P_{r0} = \left(1-\frac{e^{-(h_0+h_c)\tau}\left(1-e^{-(h_0+h_c)\tau_a}\right)}{(h_0+h_c)\tau_a}\right)\cdot\frac{h_0}{h_0+h_c} = 0.6582823659$$

$P_{r1} = 0.5388827658$

$$\frac{(z_{1-\alpha}-z_\beta)^2}{(\log\lambda)^2} = \frac{1}{\frac{1}{d_1}+\frac{1}{d_2}} = \frac{1}{\frac{d_1+d_2}{d_1 d_2}} = \frac{d_1 d_2}{d_1+d_2}$$

$$z_\beta = z_{1-\alpha} - \log\lambda \Big/ \sqrt{\left(\frac{1}{d_1}+\frac{1}{d_2}\right)}$$

$$= \frac{\frac{N^2}{4}P_{r1}P_{r2}}{\frac{N}{2}(P_{r1}+P_{r2})} = \frac{N}{2}\frac{P_{r1}P_{r2}}{(P_{r1}+P_{r2})}$$

Homework 5

1.96 , -1.28

$$d = 4\left(\frac{z_\alpha - z_\beta}{\log\lambda}\right)^2 = 370.8954808$$

1. A study design has been discussed in a study team for treating B-cell lymphoma in the second line patient population. The study will randomize subjects in 1:1 ratio

- After thorough literature search, the study team would like to assume 20 months for the median survival time in the standard of care
- The expected median survival time in the new treatment arm is 28 months
- The enrollment period is 18 months
- The minimum follow-up time for each subject is 24 months

$P_{r1} = 0.554547071$

$P_{r2} = 0.676161808$

$$\frac{d_1}{P_{r1}} = \frac{d_2}{P_{r2}} = \frac{N}{2} \qquad d_1 = \frac{N}{2}P_{r1} \qquad d_2 = \frac{N}{2}P_{r2}$$

$N_1 = 608.67$

How many events will be needed to reach 90% power at the 1-sided significant level of 0.025?

How many subjects should be planned? 625.848318 312.9 ≈ 313

What is the number of subjects if more investigate sites are available and the enrollment period is shortened to 12 months?

What would be the power loss if the hazard ratio is 1 during the first 4 months of treatment?

Please add your own assumption on the rates of loss of follow-up and re-answer the questions above.

2. The study team learned from clinicaltrial.gov that a competitor's trial for the same indication requires only 300 subjects and will take only 3 years to complete. Discuss what you think of the competitor's trial design.

3. Let $T_i \sim \exp(h_i)$, $T_j \sim \exp(h_j)$, and $T_i \perp T_j$. Show $P(T_i \geq T_j) = \frac{h_i}{h_i+h_j}$.

$$= \int_0^\infty \int_0^v f_j(u) f_i(v)\, du\, dv$$

$$= \int_0^\infty$$

$$f_i(t) = h_i e^{-h_i t} \qquad S_i(t) = P_i(T_i \geq t) = 1 - F_i(t) = e^{-h_i t}$$

$$f_j(t) = h_j e^{-h_j t} \qquad = e^{-h_j t}$$

$$p(T_i \geq T_j) = 1 - P(T_j < T_i)$$

$$= 1 - \left(\int_0^\infty P(T_j < T_i \mid T_i = t) f_i(t)\, dt\right)$$