Survival Analysis

Statistics in Medical Research Fall Series

Marcio Augusto Diniz, Ph.D.
Biostatistics and Bionformatics Research Center
Cedars Sinai Medical Center

November 1, 2017

Summary

1. Introduction
2. Graphical representation
3. Comparing survival curves
4. Regression models

Clinical trial

- Double-blind study;
- Groups: A (new treatment) and B (standard treatment);
- Aim: Comparing the percent of remission between groups along of three years.

Example

Relapse of autoimmune hepatitis

Data

Treatment	Relapse	No Relapse	Drop out	Total
A	11	12	8	31
B	20	3	7	30

- p: percentage of success (no relapse) considering ITT principle;
- $p_A = 38.7\%$ (12/31) and $p_B = 10\%$ (3/30);
- $H_0 : p_A = p_B$ vs $H_1 : p_A \neq p_B$, p value = 0.015.

Is time to remission important?

Relapse of autoimmune hepatitis

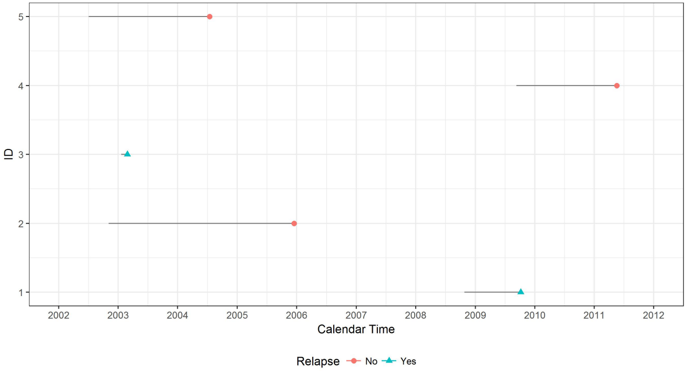

Figure: Follow up of the first 5 patients by calendar time

Is time to remission important?

Relapse of autoimmune hepatitis

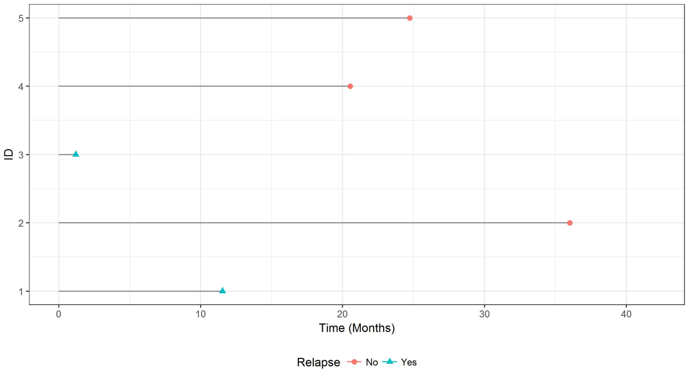

Figure: Follow up of the first 5 patients by trial time

Is time to remission important?

Relapse of autoimmune hepatitis

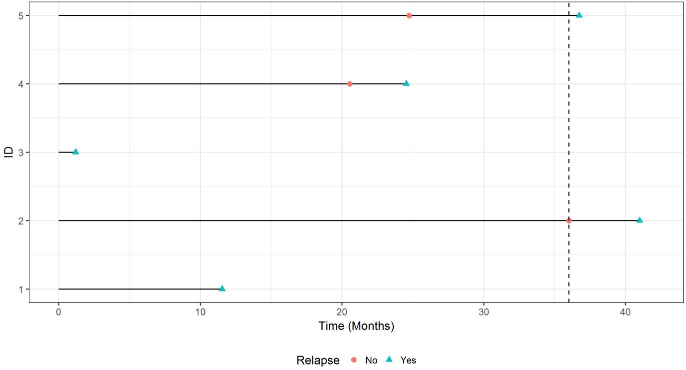

Figure: Hypothetical follow up of the first 5 patients until relapse by trial time

Definitions

- Start time;
- Length of follow-up;
- Clinical Endpoint (Death, Relapse);

Definitions

- Start time;
- Length of follow-up;
- Clinical Endpoint (Death, Relapse);

Challenges

- Loss of patient follow-up.

What is censoring?

- It is any event that does not allow us to observe our endpoint;
- It should not be excluded.

What is censoring?

- It is any event that does not allow us to observe our endpoint;
- It should not be excluded.

Assumptions

- It is not informative: censored patients would have the same probability of experiencing a event as non-censored patients.

Summary

- It is a methodology to estimate survival curves considering censoring;

Time	n at risk	n event	survival
0	61	0	1
36	60	1	$1 - 1/60$
56	58	1	$1 - 1/60 \times 1/58$

Kaplan-Meier curves

Relapse of autoimmune hepatitis

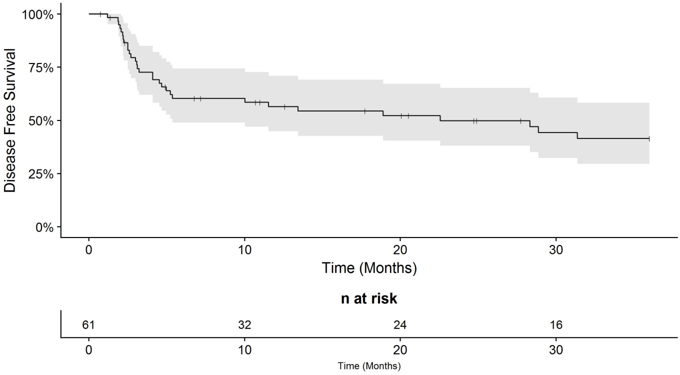

Figure: Disease free survival combining treatments A and B

Disease free survival at specific times

- 1 year: 56.48%, 95% CI: [44.94 ; 70.98]
- 2 years: 47.07%, 95% CI: [35.17 ; 63]

Kaplan-Meier curves

Relapse of autoimmune hepatitis

Disease free survival at specific times

- 1 year: 56.48%, 95% CI: [44.94 ; 70.98]
- 2 years: 47.07%, 95% CI: [35.17 ; 63]

Median disease free survival

- It represents the time such that 50% had experienced the event of interest; In this case, 22.57 months;
- It is usually calculated using the inverse of the Kaplan-Meyer curves;
- It is not always possible to calculate.

Kaplan-Meier curves

Relapse of autoimmune hepatitis

Disease free survival at specific times

- 1 year: 56.48%, 95% CI: [44.94 ; 70.98]
- 2 years: 47.07%, 95% CI: [35.17 ; 63]

Mean disease free survival

- It is the area under the Kaplan-Meier curve;
- If there is censoring then the mean survival is not a good summary because the area under the curve is underestimated.

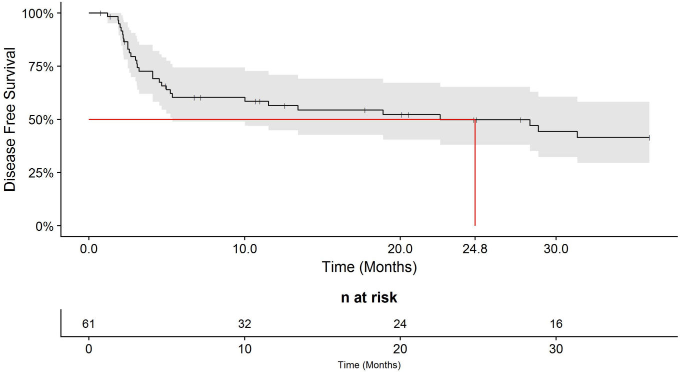

Figure: Disease free survival combining treatments A and B with median survival of 22.57 months and censoring rate of 24.5%.

Summary

Kaplan-Meier curves

Relapse of autoimmune hepatitis

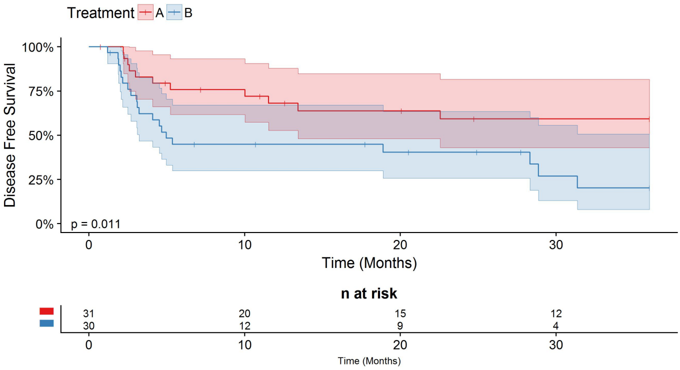

Figure: Free disease survival by treatments

Log-rank Test

- H_0: there are no differences between the treatments;
- H_1: there are differences between the treatments;
- If there are more than two curves, the test cannot indicate which curves are different from the others;
- It gives all the events the same weight.

Summary

Probabilistic model

- T: time to event of interest;
- $T \sim$ distribution of probability;
- $\lambda(t)$ is the hazard function which represents the instantaneous rate of relapse:
  - It is not a probability;
  - It is a rate of relapse at time t.

Hazard functions

Relapse of autoimmune hepatitis

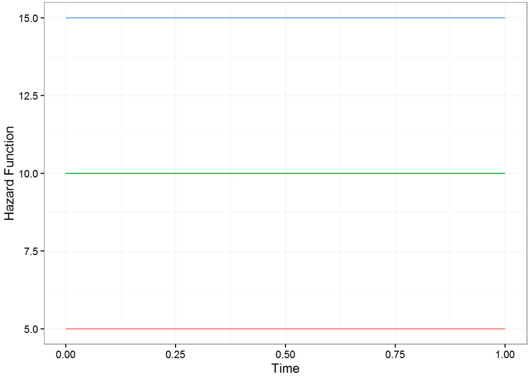

Figure: Hazard Function for Exponential distribution

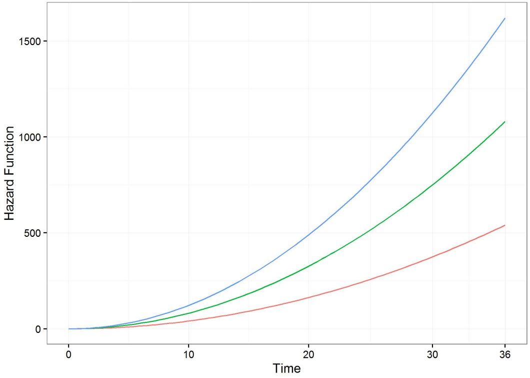

Figure: Increasing Hazard Function for Weibull distribution

Hazard functions

Relapse of autoimmune hepatitis

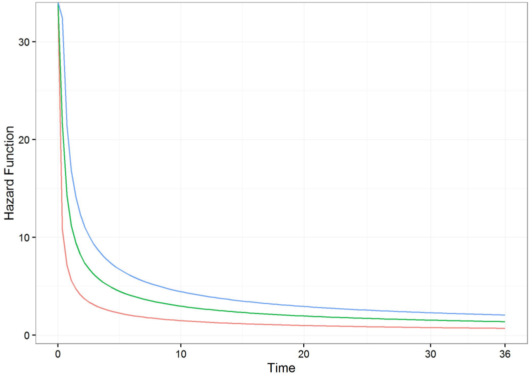

Figure: Decreasing Hazard Function for Weibull distribution

Proportional hazards model

- T: time to event of interest;
- $T \sim$ distribution of probability;
- $\lambda(t)$ is written as proportional a base hazard function,

$$\lambda_A(t) = \lambda_B(t) \times \exp\{\beta_1 \times \text{Treatment A}\}$$

- It requires the definition of a distribution of probability to define $\lambda_0(t)$.

Simple Cox proportional hazards model

- T: time to event of interest;
- $T \sim$ distribution of probability;
- $\lambda(t)$ is written as proportional a base hazard function,

$$\lambda_A(t) = \lambda_B(t) \times \exp\{\beta_1 \times \text{Treatment A}\}$$

- It does not requires the definition of a distribution of probability to define $\lambda_B(t)$.
- If $\beta_1 = 0$, then $\lambda_A(t) = \lambda_B(t)$.

Coefficients	Estimate	Std. Error	z value	p value
(Treatment A) β_1	-0.92	0.37	2.45	0.014

Table: Fitted Simple Cox regression

What does this p value mean?

- $H_0 : \beta_1 = 0 \qquad H_1 : \beta_1 \neq 0.$

Coefficients	Estimate	Std. Error	z value	p value
(Treatment A) β_1	-0.92	0.37	2.45	0.014

Table: Fitted Simple Cox regression

How to interpret the coefficients?

- We calculate the hazard ratio,

$$HR(relapse|A:B) = \exp\{\beta_1\} = 0.39$$

Coefficients	Estimate	Std. Error	z value	p value
(Treatment A) β_1	-0.92	0.37	2.45	0.014

Table: Fitted Simple Cox regression

How to interpret the coefficients?

- We calculate the hazard ratio,

$$HR(relapse|A:B) = \exp\{\beta_1\} = 0.39$$

- The treatment A has a hazard of relapse 60% ($100 \times (1 - 0.39)$) lower than treatment A;

Cox Regression

Relapse of autoimmune hepatitis

Coefficients	Estimate	Std. Error	z value	p value
(Treatment A) β_1	-0.92	0.37	2.45	0.014

Table: Fitted Simple Cox regression

How to interpret the coefficients?

- We calculate the hazard ratio,

$$HR(relapse|A:B) = \exp\{\beta_1\} = 0.39$$

- The treatment B has a hazard of relapse 2.53 (1/0.39) (95% CI: 1.20 ; 5.31) times higher than the treatment A.

Proportional hazards assumption

- It should always be verified to validate our inferences;
- If the Kaplan-Meier curves cross each other then there is evidence that the assumption is not verified;
- However, it should be verified by a statistical test using Schoenfeld residuals.

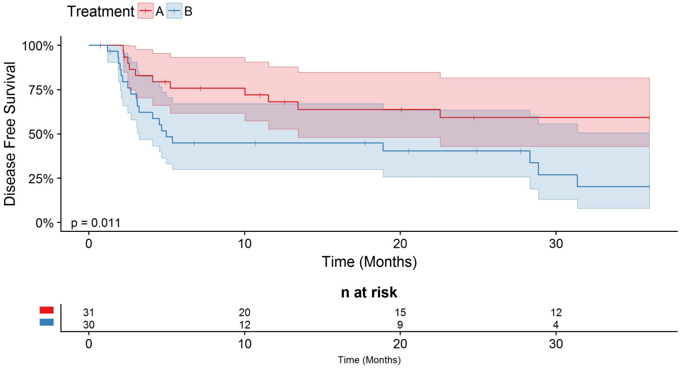

Figure: H_0: proportional hazards vs H_1: non-proportional hazards, p value $=$ 0.436

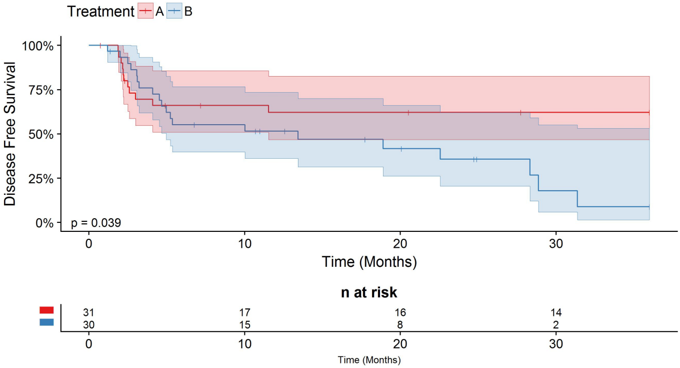

Figure: H_0: proportional hazards vs H_1: non-proportional hazards, p value $=$ 0.004

Cox Regression

Relapse of autoimmune hepatitis

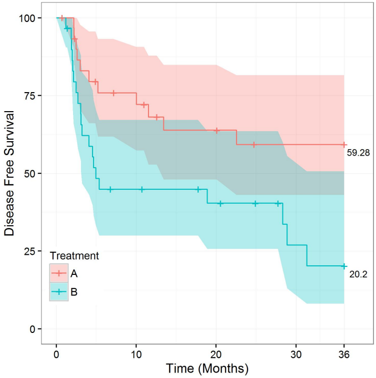

Figure: Estimated survival curves by Kaplan-Meier

Cox Regression

Relapse of autoimmune hepatitis

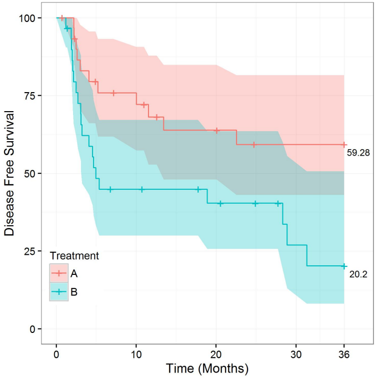

Figure: Estimated survival curves by Kaplan-Meier

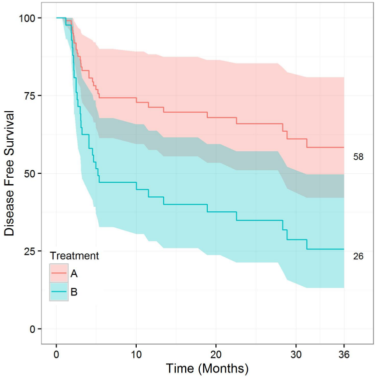

Figure: Estimated survival curves by Cox regression

Multivariable Cox proportional hazards model

- T: time to event of interest;
- $T \sim$ distribution of probability;
- $\lambda(t)$ is written as proportional a base hazard function,

$$\begin{aligned}\lambda_{A,Yes}(t) &= \lambda_{B,No}(t) \times \\ &\exp\{\beta_1 \text{Treatment: A} + \beta_2 \text{Anti-SLA: Yes}\}\end{aligned}$$

Multivariable Cox proportional hazards model

- T: time to event of interest;
- $T \sim$ distribution of probability;
- $\lambda(t)$ is written as proportional a base hazard function,

$$\begin{aligned} \lambda_{A,Yes}(t) &= \lambda_{B,No}(t) \times \\ &\quad \exp\{\beta_1 \text{Treatment: A} + \beta_2 \text{Anti-SLA: Yes}\} \end{aligned}$$

- $\beta_1 = 0$ implies to
  - $\lambda_{A,Yes}(t) = \lambda_{B,Yes}(t)$
  - $\lambda_{A,No}(t) = \lambda_{B,No}(t)$;

Multivariable Cox proportional hazards model

- T: time to event of interest;
- $T \sim$ distribution of probability;
- $\lambda(t)$ is written as proportional a base hazard function,

$$\lambda_{A,Yes}(t) = \lambda_{B,No}(t) \times \exp\{\beta_1 \text{Treatment: A} + \beta_2 \text{Anti-SLA: Yes}\}$$

- $\beta_2 = 0$ implies to
  - $\lambda_{A,Yes}(t) = \lambda_{A,No}(t)$
  - $\lambda_{B,Yes}(t) = \lambda_{B,No}(t)$.

Coefficients	Estimate	Std. Error	z value	p value
(Treatment A) β_1	-0.87	0.38	2.32	0.021
(Anti-SLA: Yes) β_2	1.11	0.38	2.89	0.003

Table: Fitted Multivariable Cox regression

What does these p values mean?

- $H_0 : \beta_1 = 0 \qquad H_1 : \beta_1 \neq 0.$
- $H_0 : \beta_2 = 0 \qquad H_1 : \beta_2 \neq 0.$

Coefficients	Estimate	Std. Error	z value	p value
(Treatment A) β_1	-0.87	0.38	2.32	0.021
(Anti-SLA: Yes) β_2	1.11	0.38	2.89	0.003

Table: Fitted Multivariable Cox regression

How to interpret the coefficients?

- We calculate the hazard ratio,

$$HR(relapse|A : B) = \exp\{\beta_1\} = 0.418$$

- The treatment B has a hazard of relapse 2.38 (1/0.418) (95% CI: 1.14 ; 5.05) times higher than the treatment A.

Multivariable Cox proportional hazards model

- T: time to event of interest;
- $T \sim$ distribution of probability;
- $\lambda(t)$ is written as proportional a base hazard function,

$$
\begin{aligned}
\lambda_{A,Yes}(t) &= \lambda_{B,No}(t) \times \\
&\exp\{\beta_1 \text{Treatment: A} + \beta_2 \text{Anti-SLA: Yes} + \\
&\beta_3 \text{Treatment: A} \times \text{Anti-SLA: Yes}\}
\end{aligned}
$$

Multivariable Cox proportional hazards model

- T: time to event of interest;
- $T \sim$ distribution of probability;
- $\lambda(t)$ is written as proportional a base hazard function,

$$\begin{aligned}\lambda_{A,Yes}(t) &= \lambda_{B,No}(t) \times \\ &\exp\{\beta_1 \text{Treatment: A} + \beta_2 \text{Anti-SLA: Yes} + \\ &\beta_3 \text{Treatment: A} \times \text{Anti-SLA: Yes}\}\end{aligned}$$

- $\beta_3 = 0$ implies to
  - $\lambda_{A,Yes}(t) - \lambda_{B,Yes}(t) = \lambda_{A,No}(t) - \lambda_{B,No}(t)$

Multivariable Cox proportional hazards model

- T: time to event of interest;
- $T \sim$ distribution of probability;
- $\lambda(t)$ is written as proportional a base hazard function,

$$\begin{aligned}\lambda_{A,Yes}(t) &= \lambda_{B,No}(t) \times \\ &\exp\{\beta_1 \text{Treatment: A} + \beta_2 \text{Anti-SLA: Yes} + \\ &\beta_3 \text{Treatment: A} \times \text{Anti-SLA: Yes}\}\end{aligned}$$

- $\beta_3 = 0$ and $\beta_1 = 0$ implies to
  - $\lambda_{A,Yes}(t) = \lambda_{B,Yes}(t)$
  - $\lambda_{A,No}(t) = \lambda_{B,No}(t)$;

Multivariable Cox proportional hazards model

- T: time to event of interest;
- $T \sim$ distribution of probability;
- $\lambda(t)$ is written as proportional a base hazard function,

$$\begin{aligned}\lambda_{A,Yes}(t) &= \lambda_{B,No}(t) \times \\ &\exp\{\beta_1 \text{Treatment: A} + \beta_2 \text{Anti-SLA: Yes} + \\ &\beta_3 \text{Treatment: A} \times \text{Anti-SLA: Yes}\}\end{aligned}$$

- $\beta_3 = 0$ and $\beta_2 = 0$ implies to
  - $\lambda_{A,Yes}(t) = \lambda_{A,No}(t)$
  - $\lambda_{B,Yes}(t) = \lambda_{B,No}(t)$.

References

- British Medical Journal
- Clark, T. G., Bradburn, M. J., Love, S. B., & Altman, D. G. (2003). Survival analysis part I: basic concepts and first analyses. British journal of cancer, 89(2), 232.
- Bradburn, M. J., Clark, T. G., Love, S. B., & Altman, D. G. (2003). Survival analysis part II: multivariate data analysis - an introduction to concepts and methods. British journal of cancer, 89(3), 431.

References

- Bradburn, M. J., Clark, T. G., Love, S. B., & Altman, D. G. (2003). Survival analysis Part III: multivariate data analysis-choosing a model and assessing its adequacy and fit. British journal of cancer, 89(4), 605.
- Clark, T. G., Bradburn, M. J., Love, S. B., & Altman, D. G. (2003). Survival analysis part IV: further concepts and methods in survival analysis. British journal of cancer, 89(5), 781.
- Singh R, Mukhopadhyay K. Survival analysis in clinical trials: Basics and must know areas. Perspectives in clinical research. 2011 Oct 1;2(4):145.