Research Lessons from Survival Analysis

Joshua Naranjo
Statistical Consulting Center
Department of Statistics
Western Michigan University
A modeling framework

Common elements

1. Data
2. Model
3. Fit criterion

Data ← Model → Fit Criterion
Ex. The two-sample problem

1. **Data:** \( \{x_i\}, \{y_j\} \)

2. **Model:** \( X \sim N(\mu_1, \sigma^2), \ Y \sim N(\mu_2, \sigma^2) \)

3. **Fit criterion:** Likelihood

---

\[
\text{Data Model Fit Criterion}
\]

\[
\overset{\text{Model}}{X \sim N(\bar{x}, s_p^2), \ Y \sim N(\bar{y}, s_p^2)}
\]
Ex. The two-sample problem

1. Data: \( \{x_i\}, \{y_j\} \)

2. Model: \( X \sim N(\mu_1, \sigma^2), \ Y \sim N(\mu_2, \sigma^2) \)

3. Fit criterion: Likelihood

\[ \begin{align*}
\text{Data} & \quad \text{Model} \\
\downarrow & \quad \uparrow \\
\text{Fit Criterion} & \\
\end{align*} \]

\[ \hat{\text{Model}} : \ X \sim N(\bar{x}, s_p^2), \ Y \sim N(\bar{y}, s_p^2) \]

- Test: \( \frac{\bar{y} - \bar{x}}{s_p \sqrt{1/n_1 + 1/n_2}} > t_\alpha \)

- Estimate: \( \bar{y} - \bar{x} \quad \left( \text{SE} = s_p \sqrt{1/n_1 + 1/n_2} \right) \)
1. **Data:** \( \{x_i\}, \{y_j\} \)

2. **Model:** \( X \sim N(\mu_1, \sigma^2_1), \ Y \sim N(\mu_2, \sigma^2_2) \)

3. **Fit criterion:** Likelihood

\[
\begin{align*}
\text{Data} & \quad \text{Model} \\
\downarrow & \quad \uparrow \\
\text{Fit Criterion} & \\
\end{align*}
\]

**Model:** \( X \sim N(\mu_1, s^2_1), \ Y \sim N(\mu_2, s^2_2) \)

- **Test:** \( \frac{\bar{y} - \bar{x}}{\sqrt{s^2_1/n_1 + s^2_2/n_2}} > t_\alpha \)

- **Estimate:** \( \bar{y} - \bar{x} \left( SE = \sqrt{s^2_1/n_1 + s^2_2/n_2} \right) \)
Wilcoxon Rank-Sum

1. Data: \( \{x_i\}, \{y_j\} \)
2. Model: \( X \sim F(t), \ Y \sim F(t - \Delta) \)
3. Fit criterion: Rank dispersion

\[ \text{Data} \quad \xleftarrow{} \quad \text{Model} \quad \xrightarrow{} \quad \text{Fit Criterion} \]

Model: \( X \sim F(t), \ Y \sim F(t - \hat{\Delta}), \) where \( \hat{\Delta} = \text{med} \{y_j - x_i\} \)

- Test: \( \frac{\sum R_i - E_0}{\sqrt{V_0}} > z_\alpha \)
- Estimate: \( \hat{\Delta} = \text{med} \{y_j - x_i\} \)
Evaluation

Data Model

Fit Criterion

Pros and cons

1. Efficiency
   - Null (size, etc.)
   - Alternative (power under various configurations)

2. Robustness
   - Broadness of assumptions
   - Sensitivity to violations

3. Simplicity and interpretability

4. Computing complexity
Example: The two-sample problem

<table>
<thead>
<tr>
<th></th>
<th>Student’s t</th>
<th>Welch t</th>
<th>Rank Sum</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Efficiency</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>2. Robustness</td>
<td>x</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>3. Interpret</td>
<td>✓</td>
<td>✓</td>
<td>x</td>
</tr>
<tr>
<td>4. Computing</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
</tbody>
</table>
More data structures

- The $k$-sample problem ($Y_1, Y_2, Y_3$)
- Regression ($X_1, X_2, X_3, Y$)
- Correlation ($X, Y$)
- Categorical data ($f_x, f_y$)
- Survival data ($T_x, T_y$)
- Longitudinal data ($\Sigma_y$)
- Time series data ($Y_t$)
- Mixture data ($pY_1 + (1 - p)Y_2$)
- ...
# Two-sample vs Two-sample

Age at death (from all causes)

<table>
<thead>
<tr>
<th>Study 1 (Estrogen)</th>
<th>Study 2 (Vitamin E)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Treatment (n = 1680)</td>
<td>Control (n = 1722)</td>
</tr>
<tr>
<td>66</td>
<td>75</td>
</tr>
<tr>
<td>68</td>
<td>58</td>
</tr>
<tr>
<td>72</td>
<td>79</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>51</td>
</tr>
<tr>
<td>44</td>
<td></td>
</tr>
</tbody>
</table>

Joshua Naranjo  
Research Lessons from Survival Analysis
Crucial difference

1. Study 1 is randomized
2. Study 2 is not randomized
   - Treatment group has higher education
   - Treatment group more female

Need to adjust for education and gender

⇒ Two-sample Regression
The two-sample problem

- Response is time-to-event (e.g. survival time)
- Patients enter the study serially (e.g. patients are referred to study upon diagnosis)
- Two groups of patients (e.g. randomized treatment arms)
<table>
<thead>
<tr>
<th>subj</th>
<th>days</th>
<th>status</th>
<th>group</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>8</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>2</td>
<td>23</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>3</td>
<td>23</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>4</td>
<td>52</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>5</td>
<td>60</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>74</td>
<td>828</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>75</td>
<td>976</td>
<td>0</td>
<td>2</td>
</tr>
</tbody>
</table>

days = time from randomization to death or censoring

status = \[
\begin{align*}
1 & \quad \text{if death} \\
0 & \quad \text{if censored (e.g. subject 4 enrolled 52 days ago)}
\end{align*}
\]
group = \[
\begin{align*}
1 & \quad \text{if treatment 1} \\
2 & \quad \text{if treatment 2}
\end{align*}
\]
### Survival data

<table>
<thead>
<tr>
<th>subj</th>
<th>days</th>
<th>group</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>8</td>
<td>2</td>
</tr>
<tr>
<td>2</td>
<td>23</td>
<td>1</td>
</tr>
<tr>
<td>3</td>
<td>23</td>
<td>2</td>
</tr>
<tr>
<td>4</td>
<td>&gt;52</td>
<td>2</td>
</tr>
<tr>
<td>5</td>
<td>60</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>74</td>
<td>828</td>
<td>1</td>
</tr>
<tr>
<td>75</td>
<td>&gt;976</td>
<td>2</td>
</tr>
</tbody>
</table>

- **days** = time from randomization to death or censoring
- **status** = \[ \begin{cases} 1 & \text{if death} \\ 0 & \text{if censored (e.g. subject 4 enrolled 52 days ago)} \end{cases} \]
- **group** = \[ \begin{cases} 1 & \text{if treatment 1} \\ 2 & \text{if treatment 2} \end{cases} \]
Q: How do we accommodate censoring? (L1)

<table>
<thead>
<tr>
<th>subj</th>
<th>days</th>
<th>group</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>8</td>
<td>2</td>
</tr>
<tr>
<td>2</td>
<td>23</td>
<td>1</td>
</tr>
<tr>
<td>3</td>
<td>23</td>
<td>2</td>
</tr>
<tr>
<td>4</td>
<td>&gt;52</td>
<td>2</td>
</tr>
<tr>
<td>5</td>
<td>60</td>
<td>1</td>
</tr>
<tr>
<td>:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>74</td>
<td>828</td>
<td>1</td>
</tr>
<tr>
<td>75</td>
<td>&gt;976</td>
<td>2</td>
</tr>
</tbody>
</table>

The KM estimate of \( S(t) = P[T > t] \):

\[
\hat{S}(0) = 1 \\
\hat{S}(8) = \frac{74}{75} = 0.9867 \\
\hat{S}(23) = \left(\frac{74}{75}\right)\left(\frac{72}{74}\right) = 0.9600 \\
\hat{S}(60) = \left(\frac{74}{75}\right)\left(\frac{72}{74}\right)\left(\frac{70}{71}\right) = 0.9465
\]
Kaplan-Meier Curve

Survival curve showing the decrease in survival rate over time.
Test for equality of groups

\[ H_0 : S_1(\cdot) = S_2(\cdot) \text{ vs } H_1 : \text{Not} \]

Two early approaches

1. Gehan (1965)
   Adapt rank tests to accommodate censoring

2. Mantel (1968)
   Adapt frequency tables to accommodate censoring (L2)
Adapt rank tests to accommodate censoring

The Wilcoxon Rank-Sum Test (1945)

<table>
<thead>
<tr>
<th>Sample A</th>
<th>Sample B</th>
</tr>
</thead>
<tbody>
<tr>
<td>Percent kill</td>
<td>Rank</td>
</tr>
<tr>
<td>68</td>
<td>12.5</td>
</tr>
<tr>
<td>68</td>
<td>12.5</td>
</tr>
<tr>
<td>59</td>
<td>3</td>
</tr>
<tr>
<td>72</td>
<td>15</td>
</tr>
<tr>
<td>64</td>
<td>8</td>
</tr>
<tr>
<td>67</td>
<td>10</td>
</tr>
<tr>
<td>70</td>
<td>14</td>
</tr>
<tr>
<td>74</td>
<td>16</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>542</strong></td>
</tr>
</tbody>
</table>
Adapt rank tests to accommodate censoring

Idea: Add up the ranks of the Group 1 subjects

<table>
<thead>
<tr>
<th>subj</th>
<th>days</th>
<th>group</th>
<th>Rank</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>8</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>23</td>
<td>1</td>
<td>2.5</td>
</tr>
<tr>
<td>3</td>
<td>23</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>&gt;52</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>60</td>
<td>1</td>
<td>?</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>74</td>
<td>828</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>75</td>
<td>&gt;976</td>
<td>2</td>
<td>?</td>
</tr>
</tbody>
</table>
Adapt rank tests to accommodate censoring

Idea: Add up the ranks of the Group 1 subjects

<table>
<thead>
<tr>
<th>group 1</th>
<th>group 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>days</td>
<td>rank</td>
</tr>
<tr>
<td>23</td>
<td>2.5</td>
</tr>
<tr>
<td>60</td>
<td>?</td>
</tr>
<tr>
<td>&gt;52</td>
<td></td>
</tr>
</tbody>
</table>

828 | 74 |
\_
?   | ？
Test for equality of groups

The Mann-Whitney version of Wilcoxon Rank-Sum test

\[ W = \sum_i \sum_j U_{ij}, \]

where \( U_{ij} = \begin{cases} 
-1 & x_i < y_j \\
0 & x_i = y_j \\
1 & x_i > y_j 
\end{cases} \)
The Mann-Whitney version of Wilcoxon Rank-Sum test

\[ W = \sum_i \sum_j U_{ij}, \]

where \( U_{ij} = \begin{cases} 
-1 & x_i < y_j \\
0 & x_i = y_j \\
1 & x_i > y_j 
\end{cases} \)

Let

\[ U^*_{ij} = \begin{cases} 
-1 & x_i < y_j \text{ or } x_i < y^*_j \\
0 & x_i = y_j \text{ or } x_i^* < y_j \text{ or } x_i > y^*_j \\
1 & x_i > y_j \text{ or } x_i^* > y_j 
\end{cases} \]

"Gehan (1965) generalization of Wilcoxon"
Test for equality of groups

"Efron (1967) modification of Gehan-Wilcoxon"

<table>
<thead>
<tr>
<th>subj</th>
<th>days</th>
<th>status</th>
<th>group</th>
<th>Rank</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>8</td>
<td>1</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>23</td>
<td>1</td>
<td>1</td>
<td>2.5</td>
</tr>
<tr>
<td>3</td>
<td>23</td>
<td>1</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>52</td>
<td>0</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>60</td>
<td>1</td>
<td>1</td>
<td>?</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>74</td>
<td>828</td>
<td>1</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>75</td>
<td>976</td>
<td>0</td>
<td>2</td>
<td></td>
</tr>
</tbody>
</table>

While it is unlikely that $60 > 52^*$, it is more likely that $828 > 52^*$, but Gehan treats them as equally uninformative ($U_{ij}^* = 0$).

$$U_{ij}^* = 1 - 2\frac{\hat{S}(x_i)}{\hat{S}(y_j)} \text{ when } x_i > y_j^*$$
Adapt frequency tables to accommodate censoring

Recall the $\chi^2$ test

<table>
<thead>
<tr>
<th></th>
<th>Yes</th>
<th>No</th>
</tr>
</thead>
<tbody>
<tr>
<td>Grp 1</td>
<td>a</td>
<td>b</td>
</tr>
<tr>
<td>Grp 2</td>
<td>c</td>
<td>d</td>
</tr>
</tbody>
</table>

Then

$$\chi^2 = \frac{(a - E)^2}{V}$$

where

$$E = \frac{n_1 m_1}{n} \quad \text{and} \quad V = \frac{n_1 n_2 m_1 m_2}{n^2(n - 1)}$$
Test for equality of groups

"Cochran-Mantel-Haenszel test (1959)"

When data are stratified according to control variables (e.g. race or gender)

\[
\begin{array}{c|c|c|c|c|c}
& \text{Yes} & \text{No} & \text{Yes} & \text{No} & \text{Yes} & \text{No} \\
\hline
\text{Grp 1} & a_1 & b_1 & a_2 & b_2 & \cdots & a_K & b_K \\
\text{Grp 2} & c_1 & d_1 & c_2 & d_2 & & c_K & d_K \\
\end{array}
\]

\[
\chi^2_{CMH} = \frac{\left( \sum a_i - \sum E_i \right)^2}{\sum V_i}
\]

Q: Why not \( \sum \left( \frac{a_i - E_i}{V_i} \right)^2 \)

Be cynical. (L3)
Test for equality of groups

“Logrank test” (Mantel, 1966)

Day 8: 1 death, and it came from Group 2

\[
\begin{array}{c|ccc}
\text{Grp 1} & 0 & 40 & 40 \\
\text{Grp 2} & 1 & 34 & 35 \\
\end{array}
\Rightarrow a_1 = 0, \quad E_1 = \frac{40}{75}
\]

Day 23: 2 deaths, one from each group

\[
\begin{array}{c|ccc}
\text{Grp 1} & 1 & 39 & 40 \\
\text{Grp 2} & 1 & 33 & 34 \\
\end{array}
\Rightarrow a_2 = 1, \quad E_2 = 2 \left(\frac{40}{74}\right)
\]

Day 60: 1 death, and it came from Group 1

\[
\begin{array}{c|ccc}
\text{Grp 1} & 1 & 38 & 39 \\
\text{Grp 2} & 0 & 32 & 32 \\
\end{array}
\Rightarrow a_3 = 1, \quad E_3 = \frac{39}{71}
\]

\[Z = \frac{\sum (a_j - E_j)}{\sqrt{\sum V_j}}, \text{ sum over distinct event times}\]
Test for equality of groups

Pros and cons? Not obvious.

While the Mantel logrank test is

\[ Z_L = \frac{\sum (a_j - E_j)}{\sqrt{\sum V_j}} \]

the Gehan-Wilcoxon can be written as

\[ Z_W = \frac{\sum n_j(a_j - E_j)}{\sqrt{\sum n_j^2 V_j}} \]

where \( n_j \) is the total number of subjects at risk at event time \( j \).

⇒ The Wilcoxon gives more weight to early differences
## Example: Survival times for heart transplant patients

<table>
<thead>
<tr>
<th>Subj</th>
<th>Survive</th>
<th>Trans</th>
<th>Prior</th>
<th>Age</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>5</td>
<td>0</td>
<td>0</td>
<td>52</td>
</tr>
<tr>
<td>2</td>
<td>15</td>
<td>1</td>
<td>0</td>
<td>55</td>
</tr>
<tr>
<td>3</td>
<td>37</td>
<td>1</td>
<td>0</td>
<td>41</td>
</tr>
<tr>
<td>4</td>
<td>48</td>
<td>0</td>
<td>1</td>
<td>31</td>
</tr>
<tr>
<td>:</td>
<td>:</td>
<td>:</td>
<td>:</td>
<td>:</td>
</tr>
<tr>
<td>44</td>
<td>541</td>
<td>1</td>
<td>0</td>
<td>46</td>
</tr>
<tr>
<td>45</td>
<td>670</td>
<td>1</td>
<td>0</td>
<td>48</td>
</tr>
</tbody>
</table>
Suppose that survival time $T$ has probability density function $f(t)$. The cumulative density function is

$$F(t) = P[T \leq t] = \int_0^t f(u) \, du$$

The survivor function is

$$S(t) = P[T > t] = 1 - F(t)$$

Define the hazard function as

$$h(t) = \frac{f(t)}{S(t)}$$

It can be shown that

$$f(t) = h(t)e^{-\int_0^t h(u) \, du}$$
Some hazard models:

\[
h(t) = \begin{cases} \lambda, & \text{Exponential} \\ \lambda \gamma^t, & \text{Gompertz} \\ \lambda t^\alpha, & \text{Weibull} \end{cases}
\]

Incorporating covariates like transplant, prior surgery, and age?

\[
h(t) = \begin{cases} \lambda e^{\beta_1 x_1 + \cdots + \beta_k x_k}, & \text{Exponential} \\ \lambda \gamma^t e^{\beta_1 x_1 + \cdots + \beta_k x_k}, & \text{Gompertz} \\ \lambda t^\alpha e^{\beta_1 x_1 + \cdots + \beta_k x_k}, & \text{Weibull} \end{cases}
\]

\[= \lambda_0(t) e^{\beta'x}\]

*Proportional hazards:* covariates are multiplicatively related to the hazard
Estimate parameters by maximum likelihood

\[ L = \prod_{i=1}^{n} f_i(t_i) \]

where

\[ f_i(t) = h_i(t) e^{- \int_{0}^{t} h_i(u) \, du} \]

\[ h_i(t) = \lambda_0(t) e^{\beta' x_i} \]

and

\[ \lambda_0(t) = \begin{cases} 
\lambda, & \text{Exponential} \\
\lambda \gamma^t, & \text{Gompertz} \\
\lambda t^\alpha, & \text{Weibull} 
\end{cases} \]

Problem: Specify \( \lambda_0(t) \)?
Cox (1972) proposed an estimation method for the $\beta$s without needing to specify $\lambda_0(t)$. Maximize the \textit{partial likelihood}

$$PL = \prod_{i=1}^{n} L_i$$

where $L_i$ is the conditional probability of failure at time $t_i$ given the number of cases at risk at time $t_i$. 
Example: Survival times for heart transplant patients

<table>
<thead>
<tr>
<th>Subj</th>
<th>Survive</th>
<th>Trans</th>
<th>Prior</th>
<th>Age</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>5</td>
<td>0</td>
<td>0</td>
<td>52</td>
</tr>
<tr>
<td>2</td>
<td>15</td>
<td>1</td>
<td>0</td>
<td>55</td>
</tr>
<tr>
<td>3</td>
<td>37</td>
<td>1</td>
<td>0</td>
<td>41</td>
</tr>
<tr>
<td>4</td>
<td>48</td>
<td>0</td>
<td>1</td>
<td>31</td>
</tr>
<tr>
<td>44</td>
<td>541</td>
<td>1</td>
<td>0</td>
<td>46</td>
</tr>
<tr>
<td>45</td>
<td>670</td>
<td>1</td>
<td>0</td>
<td>48</td>
</tr>
</tbody>
</table>
A death occurred 5 days after enrollment. What is the probability that it happens to patient 1 instead of to one of the others at-risk?

\[ L_1 = \frac{h_1(5)}{h_1(5) + \cdots + h_{45}(5)} \]

\[ L_2 = \frac{h_2(15)}{h_2(15) + \cdots + h_{45}(15)} \]

\[ \vdots \]

\[ L_{44} = \frac{h_{44}(541)}{h_{44}(541) + h_{45}(541)} \]

\[ L_{45} = 1 \]

Similar to

\[ P(A_i|B) = \frac{P(A_i \cap B)}{P(A_1 \cap B) + \cdots + P(A_n \cap B)} \]
The combination of PH assumption and partial likelihood

\[ PL = \prod_{i=1}^{n} L_i \]

allows estimation of \( \beta \) without specifying the baseline hazard!
Efficiency


<table>
<thead>
<tr>
<th>$\alpha = e^\beta$</th>
<th>1</th>
<th>2</th>
<th>4</th>
<th>8</th>
<th>16</th>
</tr>
</thead>
<tbody>
<tr>
<td>ARE from (4.11)</td>
<td>1.000</td>
<td>.901</td>
<td>.705</td>
<td>.502</td>
<td>.334</td>
</tr>
<tr>
<td>ARE under model (4.15)–(4.17)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No censoring</td>
<td>1.000</td>
<td>.982</td>
<td>.959</td>
<td>.914</td>
<td>.819</td>
</tr>
<tr>
<td>Censoring pattern 1</td>
<td>.991</td>
<td>.978</td>
<td>.950</td>
<td>.912</td>
<td>.819</td>
</tr>
<tr>
<td>Censoring pattern 2</td>
<td>.994</td>
<td>.987</td>
<td>.967</td>
<td>.915</td>
<td>.816</td>
</tr>
</tbody>
</table>

NOTE: The ARE increases under the larger model (4.15)–(4.17) for $H(t, \gamma)$. Censoring has little effect on these calculations. Pattern 1 has half of sample 1 censored at the median of the distribution for sample 0. Pattern 2 has one quarter of sample 1 censored at each quartile of the distribution for sample 0.
The PHREG Procedure

Dependent variable: Survive

Maximum Likelihood Estimates

<table>
<thead>
<tr>
<th>Var</th>
<th>DF</th>
<th>Estimate</th>
<th>Standard Error</th>
<th>Wald Chi-sq</th>
<th>Pr&gt; Chi-sq</th>
<th>Risk Ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>Trans</td>
<td>1</td>
<td>-1.708140</td>
<td>0.27860</td>
<td>37.590</td>
<td>0.0001</td>
<td>0.181</td>
</tr>
<tr>
<td>Prior</td>
<td>1</td>
<td>-0.421402</td>
<td>0.37100</td>
<td>1.290</td>
<td>0.2560</td>
<td>0.656</td>
</tr>
<tr>
<td>Age</td>
<td>1</td>
<td>0.058609</td>
<td>0.01505</td>
<td>15.163</td>
<td>0.0001</td>
<td>1.060</td>
</tr>
</tbody>
</table>
The hazard ratio for age is

\[
\frac{h(t; \text{Age} = x + 1)}{h(t; \text{Age} = x)} = \frac{\lambda_0(t)e^{-1.708x_1-0.421x_2+0.0586(x_3+1)}}{\lambda_0(t)e^{-1.708x_1-0.421x_2+0.0586x_3}} = e^{0.0586} = 1.0603
\]

so every additional year of age increases hazard of failure by 6% same for all \( t \).

(Implications and consequences of PH)
Example: Survival times for heart transplant patients

<table>
<thead>
<tr>
<th>Var</th>
<th>DF</th>
<th>Estimate</th>
<th>Error</th>
<th>Ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>Trans</td>
<td>1</td>
<td>-1.708140</td>
<td>0.27860</td>
<td>0.181</td>
</tr>
<tr>
<td>Prior</td>
<td>1</td>
<td>-0.421402</td>
<td>0.37100</td>
<td>0.656</td>
</tr>
<tr>
<td>Age</td>
<td>1</td>
<td>0.058609</td>
<td>0.01505</td>
<td>1.060</td>
</tr>
</tbody>
</table>

“Age increases hazard of death by 6%, and getting a transplant reduces hazard of death by 82%.”

While the age effect may be real, the magnitude of transplant effect is likely false.

Transplant ⇒ Late death
No Transplant ⇒ Early death
Example: Survival times for heart transplant patients

<table>
<thead>
<tr>
<th>Var</th>
<th>DF</th>
<th>Estimate</th>
<th>Error</th>
<th>Ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>Trans</td>
<td>1</td>
<td>-1.708140</td>
<td>0.27860</td>
<td>0.181</td>
</tr>
<tr>
<td>Prior</td>
<td>1</td>
<td>-0.421402</td>
<td>0.37100</td>
<td>0.656</td>
</tr>
<tr>
<td>Age</td>
<td>1</td>
<td>0.058609</td>
<td>0.01505</td>
<td>1.060</td>
</tr>
</tbody>
</table>

"Age increases hazard of death by 6%, and getting a transplant reduces hazard of death by 82%.”

While the age effect may be real, the magnitude of transplant effect is likely false.

Transplant $\Rightarrow$ Late death
No Transplant $\Leftarrow$ Early death
Figure 1. Survival of bees fed three different diets, pooled for all colonies ($n=4$) for reasons of clarity. Bees were fed with pollen and royal jelly together (P+ RJ+), pollen only (P+ RJ−) and with neither pollen nor RJ (P− RJ−).
Figure 2. Survival probability of males (⚫) and females (⚪) of *Monosteira unicostata* and lines of best fit by Weibull function (—) at six constant temperatures.
Q: How do we detect PH? (L4: Unobservables)

Two sample:

\[ h(t) = h_0(t)e^{\beta x_i} = \begin{cases} 
  h_0(t) & \text{Group 1 (} x = 0) \\
  h_0(t)\psi & \text{Group 2 (} x = 1) 
\end{cases} \]

\[ S(t) = \begin{cases} 
  \exp \left\{ \int_0^t h_0(u)du \right\} & \text{Group 1} \\
  \exp \left\{ \int_0^t h_0(u)\psi du \right\} & \text{Group 2} 
\end{cases} \]

implies

\[ S_2(t) = [S_1(t)]^\psi \]
Characterizations of Proportional Hazards

\[ S_2(t) = [S_1(t)]^\psi \]

Early separation

Late separation
Characterizations of Proportional Hazards

\[ S_2(t) = [S_1(t)]^\psi \]

<table>
<thead>
<tr>
<th>( \psi )</th>
<th>Maximum difference ([S_1(t)]^\psi - S_1(t)) achieved at ( S_1(t) ) equal to</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.10</td>
<td>0.0774</td>
</tr>
<tr>
<td>0.20</td>
<td>0.1337</td>
</tr>
<tr>
<td>0.30</td>
<td>0.1791</td>
</tr>
<tr>
<td>0.40</td>
<td>0.2172</td>
</tr>
<tr>
<td>0.50</td>
<td>0.2500</td>
</tr>
<tr>
<td>0.60</td>
<td>0.2789</td>
</tr>
<tr>
<td>0.70</td>
<td>0.3046</td>
</tr>
<tr>
<td>0.80</td>
<td>0.3277</td>
</tr>
<tr>
<td>0.90</td>
<td>0.3487</td>
</tr>
<tr>
<td>0.99</td>
<td>0.3660</td>
</tr>
</tbody>
</table>
Assume that $x_i$ is a binary variable and the only covariate. Then the score test for Cox-regression is the same as the log-rank if there are no ties.

LR test is asymptotically efficient for the two-sample problem when there is noninformative censoring.
The logrank test and Cox Regression

(a) Survival (%)
- Age < 65 years
- Age > 65 years

P = 0.008

(b) Survival (%)
- LTOT: no
- LTOT: yes

P = 0.042
Results of Cox's proportional hazards analysis for the patients with bronchiectasis

<table>
<thead>
<tr>
<th>Explanatory variables</th>
<th>Risk ratio</th>
<th>95% confidence interval</th>
<th>$P$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (&gt;65 years)</td>
<td>2.7</td>
<td>1.15–6.29</td>
<td>0.022</td>
</tr>
<tr>
<td>LTOT (yes)</td>
<td>3.12</td>
<td>1.47–6.90</td>
<td>0.003</td>
</tr>
</tbody>
</table>

LTOT, long-term oxygen therapy.
What about (L5: Extensions)

- Tied event times
- Repeated measures
  - Fixed covariates
  - Time-dependent covariates (e.g. co-meds, hypertension)
- Time-dependent coefficients (e.g. treatments that are less effective in later stages)
- Large $p$
Extend to Cox PH regression

- $R^2$
- Extra SS
- Partial plots
- Outlier diagnostics
- Mallows $C_p$
- AIC
- LASSO
- :
“Cancer immunotherapy activates the immune system to combat cancer. This poses the possibility of a delayed onset of clinical effect and violation of the PH assumption.” (Z. Xu, FDA)
“Cancer immunotherapy activates the immune system to combat cancer. This poses the possibility of a delayed onset of clinical effect and violation of the PH assumption.” (Z. Xu, FDA)
Design and Analysis of Cancer Immunotherapy Trials with Complex Survival Patterns – Topic Contributed Papers

Biopharmaceutical Section, International Chinese Statistical Association, Biometrics Section

Organizer(s): Zhenzhen Xu, FDA
Chair(s): Bifeng Ding, Amgen

8:35 AM Survival Analysis Using a 5-STAR Approach in Randomized Clinical Trials
Devan Mehrotra, Merck & Co., Inc; Rachel Marceau West, Merck & Co., Inc.

8:55 AM Designing Cancer Immunotherapy Trials with Complex Survival Patterns
Zhenzhen Xu, FDA; BIN ZHU, NIH/NCI; YONGSOEK PARK, University of Pittsburgh, Department of B

9:15 AM Robust Group Sequential Designs for Immunotherapy Trials
Pranab Ghosh, Cytel Inc.; Cyrus Mehta, Cytel

9:35 AM A Flexible Test/Estimation Coherent Approach to Evaluate the Treatment Effect of Immunotherapy Outcomes
Hajime Uno, Dana-Farber Cancer Institute; Miki Horiguchi, Kitasato University; Lu Tian, Stanford Univ Medicine
<table>
<thead>
<tr>
<th>Time</th>
<th>Title</th>
<th>Authors</th>
</tr>
</thead>
<tbody>
<tr>
<td>8:35 AM</td>
<td>Tools to Compare Restricted Mean Survival Times in Randomized Controlled Studies with Small Sample Data</td>
<td>Miki Horiguchi, Kitasato University; Hajime Uno, Dana-Farber Cancer Institute</td>
</tr>
<tr>
<td>8:50 AM</td>
<td>Survival Analysis in the Absence of Proportional Hazards: Defining the Relevant Null Hypothesis</td>
<td>Steven Snapinn, Alder Biopharmaceuticals; Qi Jiang, Seattle Genetics</td>
</tr>
<tr>
<td>9:05 AM</td>
<td>Modeling the Impact of Dose Intervention on Time-To-Event Outcomes</td>
<td>Amir Nikooienejad, Eli Lilly and Company; Yongming Qu, Eli Lilly and Company</td>
</tr>
<tr>
<td>9:20 AM</td>
<td>A Flexible Parametric Survival Model for Fitting Time-to-Event Data in Clinical Trials</td>
<td>Jason Liao, Merck &amp; Co. Inc.; Frank G Liu, Merck Sharp &amp; Dohme Inc.</td>
</tr>
<tr>
<td>9:35 AM</td>
<td>Teasing Out the Overall Survival Benefit with Adjustment for Treatment Switching to Other Therapies</td>
<td>Meijing Wu, AbbVie; Weili He, AbbVie; Qiming Liao, VIIV Health Care; Yabing Mai, AbbVie</td>
</tr>
<tr>
<td>9:50 AM</td>
<td>An Approach to Increase Power in Immuno-Oncology Trials When Non-Proportional Hazard Is Present</td>
<td>Nan Jia, Sanofi US</td>
</tr>
</tbody>
</table>
The NPH Working Group

- Collaboration initiated by FDA in 2016
- Methodical evaluation of available methods is needed
- Appropriate analysis for various patterns of non-PH
  - How to optimally test for treatment differences?
  - How to summarize nonconstant treatment effect?

- Organizations represented
  AZ, BMS, Merck, Boehringer Ingelheim, Novartis, Eli Lilly, Abbvie, Genentech, Roche, Bayer, Janssen, Takeda, Amgen, Pfizer, GSK, Celgene, Sanofi, Johnson & Johnson, FDA
Non-Proportional Hazards

Available methods

- Weighted logrank (Fleming and Harrington 1987)
- Piecewise logrank (Xu et al. 2016)
- Combination tests
  - Karrison 2016 (Max-Combo)
  - Gares et al. 2017
  - Yang and Prentice 2010
  - Logan 2008
- Restricted mean survival time (Uno et al. 2014)
- Generalized pairwise comparisons $U_{ij}$
  - Peron et al. 2018
  - Buyse 2010
- Weighted KM test (Pepe and Fleming 1989, 1991)
Weighted logrank tests

Logrank test (Mantel, 1968)

\[ L = \frac{\sum (a_j - E_j)}{\sqrt{\sum V_j}} \]

Wilcoxon (Gehan, 1965)

\[ W = \frac{\sum n_j (a_j - E_j)}{\sqrt{\sum n_j^2 V_j}} = \frac{\sum \hat{S}(t_j) (a_j - E_j)}{\sqrt{\sum \hat{S}^2(t_j) V_j}} \]

Weighted Logrank (Fleming and Harrington, 1987)

\[ FH = \frac{\sum W(t_j)(a_j - E_j)}{\sqrt{\sum W^2(t_j) V_j}} \]

where \( W(t) = \left[ \hat{S}(t) \right]^\rho \left[ 1 - \hat{S}(t) \right]^\gamma \)
Weighted logrank tests

\[ FH(\rho, \gamma) = \begin{cases} 
\text{Logrank} & (\rho, \gamma) = (0, 0) \\
\text{Wilcoxon} & (1, 0) \\
\text{Late difference} & (0, 1)
\end{cases} \]

Pros and cons

- Easy to understand and implement
- With correct choice of weight, can be much more efficient than either logrank or Wilcoxon
- Low efficiency with incorrect choice of weight
Karrison (2016): Reject $H_0: S_1 = S_2$ if

$$Z_m = \max(|FH(0, 0)|, |FH(1, 0)|, |FH(0, 1)|) > g_\alpha$$

- $g_\alpha$ is the upper $\alpha$ percentile of the null distribution of $\max(|Z_1|, |Z_2|, |Z_3|)$
- Calculating $g_\alpha$

\[
\alpha = P[Z_m > g_\alpha] = 1 - P[Z_m \leq g_\alpha] = 1 - P[|Z_1| \leq g_\alpha, |Z_2| \leq g_\alpha, |Z_3| \leq g_\alpha]
\]

- $(Z_1, Z_2, Z_3) \sim MN(0, \Sigma)$,
Max-Combo

\[ \text{FH}(0, 0) : p = .022 \]
\[ \text{FH}(1, 0) : p = .062 \]
\[ \text{FH}(0, 1) : p = .015 \]
\[ Z_m : p = .029 \]
Max-Combo

\[
\begin{align*}
FH(0, 0) & : p = .190 \\
FH(1, 0) & : p = .045 \\
FH(0, 1) & : p = .938 \\
Z_m & : p = .082
\end{align*}
\]
Lessons:

1. Accommodate (certain data features)
2. Adapt (existing methods)
3. Be cynical
4. Unobservable assumptions (diagnostics)
5. Extensions (broader applications)
6. Try combining methods (balanced properties)
Recent dissertation topics (Statistics, WMU)

- "Nonparametric Tests for Ordered Alternatives in a Two-Stage Nested Design" (Bang Le, 2019)
- "A Two-Stage Bayesian Variable Selection for Geo-Referenced Count Data" (Yuqian Shen, 2019)
- "Testing for Association in Stratified $2 \times 2$ Contingency Tables" (Eunice Ampah, 2018)
- "Statistical Models for Correlated Data" (Xiaomeng Niu, 2017)
- "Diagnostics for Choosing Stratified Logrank or Wilcoxon Tests" (Jhoanne Gatpatan, 2017)