Statistical Methods in Survival Analysis

Zhen-Huan (Kenny) Hu, MPH

Center for International Blood and Marrow Transplant Research
Medical College of Wisconsin

August 13, 2021

Outline

Survival function

Survival function Probability of a subject experiencing the event after time t:

$$S(t) = P(T > t)$$

where T denotes the time to event (which is a random variable).
Failure function Probability of a subject experiencing the event before or at time t (complement of $S(t)$ and CDF of time to event T):

$$F(t) = P(T \leq t) = 1 - S(t)$$

$$S(t) = 1 - F(t)$$

Event rate at time t (PDF of time to event T):

$$f(t) = \lim_{\Delta t \to 0} \frac{P(t \leq T < t + \Delta t)}{\Delta t} = \frac{d}{dt} F(t) = -\frac{d}{dt} S(t)$$

$$F(t) = \int_0^t f(u)du \quad S(t) = \int_t^\infty f(u)du$$

Hazard function

Hazard function Conditional event rate for a subject who has made it to time t without experiencing the event (also known as **hazard rate**):

$$\begin{aligned}\lambda(t) &= \lim_{\Delta t \to 0} \frac{P(t \leq T < t + \Delta t | T \geq t)}{\Delta t} \\ &= \lim_{\Delta t \to 0} \frac{P(t \leq T < t + \Delta t)}{\Delta t} \cdot \frac{1}{P(T \geq t)}\end{aligned}$$

If T is continuous:

$$\lambda(t) = \frac{f(t)}{S(t)} = -\frac{1}{S(t)} \cdot \frac{d}{dt} S(t) = -\frac{d}{dt} \log S(t)$$

Cumulative hazard function:

$$\Lambda(t) = \int_0^t \lambda(u) du = -\log S(t)$$

Relationship between survival and hazard functions

$$\lambda(t) = -\frac{d}{dt} \log S(t)$$
$$\Lambda(t) = -\log S(t)$$

$$S(t) = \exp\left[-\int_0^t \lambda(u)du\right] = \exp\left[-\Lambda(t)\right]$$

Cumulative incidence function

- The **cumulative incidence function** (CIF) is the probability that an event of type k occurs at or before the given time t. It can be also seen as the probability of **cause-specific** failure for event of type k.

$$F_k(t) = P(T \leq t \cap \delta = k)$$

  for $k = 1, \ldots, K$. Here δ indicates the event type.
- Used when there are multiple events that precludes each other. Events other than type k are called the **competing risks** of event k.
- Relationship with the survival and failure functions:

$$\sum_{k=1}^{K} F_k(t) = F(t) \quad S(t) + \sum_{k=1}^{K} F_k(t) = S(t) + F(t) = 1$$

Cause-specific hazard function

Cause-specific hazard function Conditional event rate for event of type k for a subject who has yet to experience any event:

$$\begin{aligned}\lambda_k(t) &= \lim_{\Delta t \to 0} \frac{P(t \leq T < t + \Delta t \cap \delta = k | T > t)}{\Delta t} \\ &= \lim_{\Delta t \to 0} \frac{P(t \leq T < t + \Delta t \cap \delta = k)}{\Delta t} \cdot \frac{1}{P(T > t)} \\ &= \frac{1}{S(t^-)} \cdot \frac{\partial F_k(t)}{\partial t}\end{aligned}$$

Relationship between cumulative incidence function and cause-specific hazard function:

$$F_k(t) = \int_0^t S(u^-)\lambda_k(u)du$$

Kaplan-Meier estimator

- **Assumption** Event and censoring are independent from each other.
- Also known as the product-limit estimator:

$$\hat{S}(t) = \prod_{t_j \leq t} \left(1 - \frac{d_j}{Y_j}\right)$$

  t_j: jth distinct event time. d_j: Number of events at time t_j. c_j: Number of censoring at time t_j. Y_j: Number of subjects still at risk of experiencing the event immediately prior to time t_j.
- Non-parametric (no assumption about the underlying probability distribution)
- When there is no censoring, the Kaplan-Meier estimator is identical to the empirical estimator of the survival function obtained simply by calculating the proportion of subjects who have not yet experienced the event by time t.

Kaplan-Meier estimator - Example

Without censoring

t_j	d_j	Y_j	$\hat{S}(t_j)$
0	0	38	1
22	1	38	$1 - 1/38 = 0.9737$
55	1	37	$0.9737 \times (1 - 1/37) = 0.9474$
74	1	36	$0.9474 \times (1 - 1/36) = 0.9211$
90	2	35	$0.9211 \times (1 - 2/35) = 0.8684$

With censoring

t_j	d_j	c_j	Y_j	$\hat{S}(t_j)$
0	0	0	38	1
22	1	0	38	$1 - 1/38 = 0.9737$
55	1	1	37	$0.9737 \times (1 - 1/37) = 0.9474$
74	0	1	35	$0.9474 \times (1 - 0/35) = 0.9474$
90	2	1	34	$0.9211 \times (1 - 2/34) = 0.8916$

Variance estimation for Kaplan-Meier estimator

Greenwood's formula[1]:

$$\hat{V}\left[\hat{S}(t)\right] = \hat{S}(t)^2 \sum_{t_j \leq t} \frac{d_j}{Y_j\left(Y_j - d_j\right)}$$

Example

t_j	d_j	c_j	Y_j	$\hat{S}(t_j)$	$\hat{V}\left[\hat{S}(t)\right]$
0	0	0	38	1	0
22	1	0	38	0.9737	$0.9737^2 \times [1/(38 \times 37)] = 0.00067$
55	1	1	37	0.9474	$0.9474^2 \times [1/(38 \times 37) + 1/(37 \times 36)] = 0.00131$
74	0	1	35	0.9474	$0.9474^2 \times [1/(38 \times 37) + 1/(37 \times 36)] = 0.00131$
90	2	1	34	0.8916	$0.8916^2 \times [1/(38 \times 37) + 1/(37 \times 36) + 2/(34 \times 32)] = 0.00262$

[1] John P. Klein et al. *Survival Analysis - Techniques for Censored and Truncated Data*. P92

Confidence interval for survival estimates

Linear transformation (not very good):

$$\hat{S}(t) - Z_{1-\frac{\alpha}{2}}\hat{V}\left[\hat{S}(t)\right]^{\frac{1}{2}} \leq \hat{S}(t) \leq \hat{S}(t) + Z_{1-\frac{\alpha}{2}}\hat{V}\left[\hat{S}(t)\right]^{\frac{1}{2}}$$

Arcsine-square root transformation (more robust when number at risk is low):

$$\sin^2\left\{\max\left\{0, \arcsin\left[\hat{S}(t)^{\frac{1}{2}}\right] - Z_{1-\frac{\alpha}{2}}\hat{\tau}(t)\right\}\right\} \leq \hat{S}(t) \leq \sin^2\left\{\min\left\{\frac{\pi}{2}, \arcsin\left[\hat{S}(t)^{\frac{1}{2}}\right] + Z_{1-\frac{\alpha}{2}}\hat{\tau}(t)\right\}\right\}$$

where $\hat{\tau}(t) = \hat{V}\left[\hat{S}(t)\right] / \left(4 \cdot \hat{S}(t)\left[1 - \hat{S}(t)\right]\right)$.
For $\alpha = 0.05$ (95% confidence interval), $Z_{0.975} = 1.96$.

Kaplan-Meier estimator - Example (SAS)

Generating Kaplan-Meier estimates

```
1 proc lifetest data = final;
2   strata trtgp;
3   time intxsurv * dead(0);
4 run;
```

Plotting Kaplan-Meier curves

```
1 ods graphics on / reset = all imagename = 'outfig' imagefmt = png;
2 proc lifetest data = final plots = (survival);
3   strata trtgp;
4   time intxsurv * dead(0);
5 run;
6 ods graphics off;
```

Nelson-Aalen estimator

- Non-parametric estimator for cumulative hazard function:

$$\hat{\Lambda}(t) = \sum_{t_j \leq t} \frac{d_j}{Y_j}$$

- Survival function can be therefore estimated by:

$$\hat{S}(t) = \exp\left[-\hat{\Lambda}(t)\right] = \exp\left[-\sum_{t_j \leq t} \frac{d_j}{Y_j}\right]$$

Code example

```
1 proc lifetest data = final nelson;
2   strata trtgp;
3   time intxsurv * dead(0);
4 run;
```

Estimation of CIF

Based on $F_k(t) = \int_0^t S(u^-)\lambda_k(u)du$, the CIF can be estimated by[2]:

$$\hat{F}_k(t) = \sum_{t_j \leq t} \hat{S}(t_{j-1}) \cdot \frac{r_j}{Y_j}$$

where $\hat{S}(t_{j-1})$ is just a Kaplan-Meier estimator at time t_{j-1}. r_j: Number of events of type k at time t_j. Y_j: Number of subjects still at risk of experiencing event of any type immediately prior to t_j.

[2] John P. Klein et al. *Survival Analysis - Techniques for Censored and Truncated Data*. P128

Variance of cumulative incidence

Using the **Delta method**[3]:

$$\hat{V}\left[\hat{F}_k(t)\right] = \sum_{t_j \le t}\left[\hat{F}_k(t) - \hat{F}_k\left(t_j\right)\right]^2 \frac{d_j}{Y_j\left(Y_j - d_j\right)} + \sum_{t_j \le t}\left[\hat{S}\left(t_{j-1}\right)\right]^2 \frac{r_j\left(Y_j - r_j\right)}{Y_j^3} - \sum_{t_j \le t} 2\left[\hat{F}_k(t) - \hat{F}_k\left(t_j\right)\right]\hat{S}\left(t_{j-1}\right)\frac{r_j}{Y_j^2}$$

r_j: Total number of events of type k at time t_j. d_j: Total number of events of any type at time t_j. Y_j: Total number of subjects still at risk of experiencing event of any type immediately prior to t_j.

[3] Melanie Pintilie *Competing Risks - A Practical Perspective*. P62. John P. Klein et al. *Survival Analysis - Techniques for Censored and Truncated Data*. P128, Eq. 4.7.2 is wrong.

Confidence interval for CIF estimates

Arcsine-square root transformation:

$$\sin^2\left\{\max\left\{0, \arcsin\left[\hat{F}_k(t)^{\frac{1}{2}}\right] - Z_{1-\frac{\alpha}{2}}\hat{\tau}(t)\right\}\right\} \leq \hat{F}_k(t) \leq \sin^2\left\{\min\left\{\frac{\pi}{2}, \arcsin\left[\hat{F}_k(t)^{\frac{1}{2}}\right] + Z_{1-\frac{\alpha}{2}}\hat{\tau}(t)\right\}\right\}$$

where $\hat{\tau}(t) = \hat{V}\left[\hat{F}_k(t)\right] / \left(4 \cdot \hat{F}_k(t)\left[1 - \hat{F}_k(t)\right]\right)$.
For $\alpha = 0.05$ (95% confidence interval), $Z_{0.975} = 1.96$.

CIF estimator - Example (SAS)

Generating CIF estimates

```
1 proc lifetest data = final;
2   strata trtgp;
3   time intxrel * relstat(0) / failcode = 1;
4 run;
```

Plotting CIF curves

```
1 ods graphics on / reset = all imagename = 'outfig' imagefmt = png;
2 proc lifetest data = final plots = (cif);
3   strata trtgp;
4   time intxrel * relstat(0) / failcode = 1;
5 run;
6 ods graphics off;
```

Hypothesis testing at a fixed point in time

Hypotheses for survival function at predetermined t_0

$$H_0 \colon S_1(t_0) = S_2(t_0) = \cdots = S_K(t_0)$$
$$H_a \colon \text{At least one of the } S_k(t_0) \text{ is different for } k = 1, \ldots, K$$

- Also referred to as **point-wise** comparison of survival/cumulative incidence functions.
- Same method for survival and cumulative incidence functions.

Hypothesis testing at a fixed point in time

The test statistic for point-wise comparison is a quadratic form[4]:

$$\chi^2 = \begin{bmatrix} \hat{\theta}_1 - \hat{\theta}_K \\ \hat{\theta}_2 - \hat{\theta}_K \\ \vdots \\ \hat{\theta}_{K-1} - \hat{\theta}_K \end{bmatrix}^\top \begin{bmatrix} \hat{V}_1 + \hat{V}_K & \hat{V}_K & \dots & \hat{V}_K \\ \hat{V}_K & \hat{V}_2 + \hat{V}_K & \dots & \hat{V}_K \\ \vdots & \vdots & \ddots & \vdots \\ \hat{V}_K & \hat{V}_K & \dots & \hat{V}_{K-1} + \hat{V}_K \end{bmatrix}^{-1} \begin{bmatrix} \hat{\theta}_1 - \hat{\theta}_K \\ \hat{\theta}_2 - \hat{\theta}_K \\ \vdots \\ \hat{\theta}_{K-1} - \hat{\theta}_K \end{bmatrix}$$

where $\hat{\theta}_k = \hat{S}_k(t_0)$ and $\hat{V}_k = \hat{V}\left[\hat{S}_k(t_0)\right]$ for $k = 1, \dots, K$.

For two-sample test, $\chi^2 = \left[\hat{S}_1(t_0) - \hat{S}_2(t_0)\right]^2 / \left[\hat{V}\left[\hat{S}_1(t_0)\right] + \hat{V}\left[\hat{S}_2(t_0)\right]\right]$.

If $\hat{\theta}_k - \hat{\theta}_K$ is approximately normally distributed for $k = 1, \dots, K-1$, χ^2 has a chi-squared distribution with $K - 1$ degree of freedom under the null hypothesis.

[4] John P. Klein et al. *Survival Analysis - Techniques for Censored and Truncated Data*. P286

Log-rank test

Hypotheses

$$H_0 : \lambda_1(t) = \lambda_2(t) = \cdots = \lambda_K(t), \forall t \leq \tau$$
$$H_a : \text{At least one of the } \lambda_k(t) \text{ is different for some } t \leq \tau$$

where τ is the largest time at which all groups have at least one subject at risk.

We can evaluate the difference between the cumulative hazard function of treatment group k and that of the pooled sample by using:

$$Z_k(\tau) = \int_0^\tau W_k(t) \left[\lambda_k(t) - \lambda(t)\right], k = 1, \ldots, K$$

where $W_k(t)$ is an arbitrary weight.
If $Z_k(\tau)$ is zero for all $k = 1, \ldots, K$, then the null hypothesis is true.

Log-rank test

$Z_k(\tau)$ can be estimated based on the Nelson-Aalen estimator:

$$\hat{Z}_k(\tau) = \sum_{t_j \leq \tau} W_k(t_j) \left[\frac{d_{jk}}{Y_{jk}} - \frac{d_j}{Y_j} \right], k = 1, \ldots, K$$

For log-rank test, $W_k(t_j) = Y_{jk}$:

$$\hat{Z}_k(\tau) = \sum_{t_j \leq \tau} \left[d_{jk} - \frac{d_j Y_{jk}}{Y_j} \right], k = 1, \ldots, K$$

The test statistic for log-rank test is given by a quadratic form:

$$\chi^2 = \left[\hat{Z}_1(\tau), \ldots, \hat{Z}_{K-1}(\tau) \right] \hat{\mathbf{\Sigma}}^{-1} \left[\hat{Z}_1(\tau), \ldots, \hat{Z}_{K-1}(\tau) \right]^\top$$

where $\hat{\mathbf{\Sigma}}$ is the covariance matrix of $\left[\hat{Z}_1(\tau), \ldots, \hat{Z}_{K-1}(\tau) \right]^\top$.

If $\left[\hat{Z}_1(\tau), \ldots, \hat{Z}_{K-1}(\tau) \right]^\top$ is approximately normally distributed, χ^2 has a chi-squared distribution with $K-1$ degree of freedom under the null hypothesis.

Stratified log-rank test

Hypotheses

$$H_0 \colon \lambda_{1s}(t) = \lambda_{2s}(t) = \cdots = \lambda_{Ks}(t), \forall t \leq \tau, s = 1, \ldots, S$$
$$H_a \colon \text{At least one of the } \lambda_{ks}(t) \text{ is different for some } t \leq \tau$$

Like unstratified log-rank test, $\hat{Z}_{ks}(\tau)$ for stratum s can be written as:

$$\hat{Z}_{ks}(\tau) = \sum_{t_j \leq \tau} \left[d_{jks} - \frac{d_{js} Y_{jks}}{Y_{js}} \right], k = 1, \ldots, K$$

The test statistic for a stratified log-rank test is given by:

$$\chi^2 = \left[\hat{Z}_1(\tau), \ldots, \hat{Z}_{K-1}(\tau) \right] \hat{\mathbf{\Sigma}}^{-1} \left[\hat{Z}_1(\tau), \ldots, \hat{Z}_{K-1}(\tau) \right]^\top$$

where $\hat{Z}_k(\tau) = \sum_{s=1}^{S} \hat{Z}_{ks}(\tau)$ for $k = 1, \ldots, K-1$.

Log-rank test - Example (SAS)

Log-rank test without stratification

```
1 proc lifetest data = final;
2   strata trtgp;
3   time intxsurv * dead(0);
4 run;
```

A log-rank test comparing `trtgp` will be performed.

Stratified log-rank test

```
1 proc lifetest data = final;
2   strata sex / group = trtgp;
3   time intxsurv * dead(0);
4 run;
```

A stratified log-rank test comparing `trtgp` will be performed, stratifying over `sex`.

Cox proportional hazards model

- The hazard function with a given risk vector $\mathbf{Z}$:

$$\lambda(t|\mathbf{Z}) = \lambda_0(t) \exp\left[\boldsymbol{\beta}^\top \mathbf{Z}\right]$$

  where $\lambda_0(t)$ is an arbitrary baseline hazard function. $\mathbf{Z}$ is a risk vector $[Z_1, \ldots, Z_p]^\top$ and $\boldsymbol{\beta}$ a coefficient vector $[\beta_1, \ldots, \beta_p]^\top$.
- Semi-parametric: $\lambda_0(t)$ is arbitrary.
- The **hazard ratio** (HR) (also called **relative risk**) of the hazard function with a risk vector $\mathbf{Z}_1$ vs. that with a risk vector $\mathbf{Z}_2$ is:

$$\frac{\lambda\left(t|\mathbf{Z}_1\right)}{\lambda\left(t|\mathbf{Z}_2\right)} = \frac{\lambda_0(t) \exp\left[\boldsymbol{\beta}^\top \mathbf{Z}_1\right]}{\lambda_0(t) \exp\left[\boldsymbol{\beta}^\top \mathbf{Z}_2\right]} = \exp\left[\boldsymbol{\beta}^\top \mathbf{Z}_1 - \boldsymbol{\beta}^\top \mathbf{Z}_2\right]$$

Single categorical variable with two categories

Sex female vs. male

- $Z_1 = 0$ for female (baseline)

$$\lambda(t|Z_1 = 0) = \lambda_0(t) \exp[\beta_1 Z_1] = \lambda_0(t)$$

- $Z_1 = 1$ for male

$$\lambda(t|Z_1 = 1) = \lambda_0(t) \exp[\beta_1 Z_1] = \lambda_0(t) \exp[\beta_1]$$

HR for male vs. female:

$$\frac{\lambda(t|Z_1 = 1)}{\lambda(t|Z_1 = 0)} = \frac{\lambda_0(t) \exp[\beta_1]}{\lambda_0(t)} = \exp[\beta_1]$$

What is the HR for female vs. male?

Single categorical variable with more than two categories

Race white vs. black vs. Asian

- $Z_1 = 0$ for white (baseline)

$$\lambda(t|Z_1 = 0) = \lambda_0(t)$$

- $Z_1 = 1$ for black

$$\lambda(t|Z_1 = 1) = \lambda_0(t) \exp[\beta_1]$$

- $Z_1 = 2$ for Asian

$$\lambda(t|Z_1 = 2) = \lambda_0(t) \exp[2\beta_1]$$

This is wrong!

If we model the data like so, how would the results look like?

Single categorical variable with more than two categories

Race white vs. black vs. Asian

- $Z_1 = 0, Z_2 = 0$ for white (baseline)

$$\lambda(t|Z_1 = 0, Z_2 = 0) = \lambda_0(t) \exp\left[\beta_1 Z_1 + \beta_2 Z_2\right] = \lambda_0(t)$$

- $Z_1 = 1, Z_2 = 0$ for black

$$\lambda(t|Z_1 = 1, Z_2 = 0) = \lambda_0(t) \exp\left[\beta_1 Z_1 + \beta_2 Z_2\right] = \lambda_0(t) \exp\left[\beta_1\right]$$

- $Z_1 = 0, Z_2 = 1$ for Asian

$$\lambda(t|Z_1 = 0, Z_2 = 1) = \lambda_0(t) \exp\left[\beta_1 Z_1 + \beta_2 Z_2\right] = \lambda_0(t) \exp\left[\beta_2\right]$$

HRs:

- Black vs. white: $\exp\left[\beta_1\right]$
- Asian vs. white: $\exp\left[\beta_2\right]$
- Asian vs. black: $\exp\left[\beta_2 - \beta_1\right]$

Multiple categorical variables

Sex female vs. male

- $Z_1 = 0$ for female (baseline)
- $Z_1 = 1$ for male

Race white vs. black vs. Asian

- $Z_2 = 0, Z_3 = 0$ for white (baseline)
- $Z_2 = 1, Z_3 = 0$ for black
- $Z_2 = 0, Z_3 = 1$ for Asian

Multiple categorical variables

HRs:

- Black vs. white among females:

$$\frac{\lambda(t|Z_1 = 0, Z_2 = 1, Z_3 = 0)}{\lambda(t|Z_1 = 0, Z_2 = 0, Z_3 = 0)} = \exp[\beta_2]$$

- Black vs. white among males:

$$\frac{\lambda(t|Z_1 = 1, Z_2 = 1, Z_3 = 0)}{\lambda(t|Z_1 = 1, Z_2 = 0, Z_3 = 0)} = \frac{\exp[\beta_1 + \beta_2]}{\exp[\beta_1]} = \exp[\beta_2]$$

- Asian vs. black among females: $\exp[\beta_3 - \beta_2]$
- Asian vs. black among males: $\exp[\beta_3 - \beta_2]$
- Asian male vs. black female:

$$\frac{\lambda(t|Z_1 = 1, Z_2 = 0, Z_3 = 1)}{\lambda(t|Z_1 = 0, Z_2 = 1, Z_3 = 0)} = \frac{\exp[\beta_1 + \beta_3]}{\exp[\beta_2]} = \exp[\beta_1 + \beta_3 - \beta_2]$$

Why the HR for black vs. white and Asian vs. black are the same among females and among males?

Multiple categorical variables with interactions

Add two additional terms $Z_4 = Z_1 \cdot Z_2$ and $Z_5 = Z_1 \cdot Z_3$.

Sex female vs. male

- $Z_1 = 0$ for female (baseline)
- $Z_1 = 1$ for male

Race white vs. black vs. Asian

- $Z_2 = 0, Z_3 = 0, Z_4 = 0, Z_5 = 0$ for white (baseline)
- $Z_2 = 1, Z_3 = 0, Z_4 = 0, Z_5 = 0$ for black female
- $Z_2 = 1, Z_3 = 0, Z_4 = 1, Z_5 = 0$ for black male
- $Z_2 = 0, Z_3 = 1, Z_4 = 0, Z_5 = 0$ for Asian female
- $Z_2 = 0, Z_3 = 1, Z_4 = 0, Z_5 = 1$ for Asian male

Multiple categorical variables with interactions

HRs:

- Black vs. white among females:

$$\frac{\lambda(t|Z_1 = 0, Z_2 = 1, Z_3 = 0, Z_4 = 0, Z_5 = 0)}{\lambda(t|Z_1 = 0, Z_2 = 0, Z_3 = 0, Z_4 = 0, Z_5 = 0)} = \exp[\beta_2]$$

- Black vs. white among males:

$$\frac{\lambda(t|Z_1 = 1, Z_2 = 1, Z_3 = 0, Z_4 = 1, Z_5 = 0)}{\lambda(t|Z_1 = 1, Z_2 = 0, Z_3 = 0, Z_4 = 0, Z_5 = 0)} = \frac{\exp[\beta_1 + \beta_2 + \beta_4]}{\exp[\beta_1]} = \exp[\beta_2 + \beta_4]$$

- Asian vs. black among females: $\exp[\beta_3 - \beta_2]$
- Asian vs. black among males: $\exp[\beta_3 + \beta_5 - \beta_2 - \beta_4]$

When $\beta_4 = 0$ and $\beta_5 = 0$, it falls back to the previous model where there is no interaction.

Cox proportional hazards model - Example (SAS)

Without interaction

```
proc phreg data = final;
  class trtgp sex race / ref = first param = ref;
  model intxsurv * dead(0) = trtgp sex race / rl;
  contrast 'Asian vs. black' race -1 1 / estimate = exp;
  contrast 'Asian M vs. black F' sex 1 race -1 1 / estimate = exp;
run;
```

With interactions

```
proc phreg data = final;
  class trtgp sex race / ref = first param = ref;
  model intxsurv * dead(0) = trtgp sex race race * sex / rl;
  contrast 'Asian F vs. black F' race -1 1 race * sex 0 0 / estimate = exp;
  contrast 'Asian M vs. black M' race -1 1 race * sex -1 1 / estimate = exp;
run;
```

Time-dependent Cox model

Normally the risk vector $\mathbf{Z}$ is predetermined at starting time (e.g., age, disease status at infusion). When there are risk factors the values of which change during the course of follow-up (e.g., onset of acute GVHD), they must be treated as **time-dependent covariates** $\mathbf{Z}(t)$[5]:

$$\lambda\left[t|\mathbf{Z}(t)\right] = \lambda_0(t)\exp\left[\boldsymbol{\beta}^\top \mathbf{Z}(t)\right]$$

Code example

```
1  proc phreg data = final;
2    class trtgp / ref = first param = ref;
3    model intxsurv * dead(0) = trtgp tagvhd / rl;
4    if agvhd = 1 and intxsurv >= intxagvhd then tagvhd = 1;
5    else tagvhd = 0;
6  run;
```

[5] John P. Klein et al. *Survival Analysis - Techniques for Censored and Truncated Data*. P296

Time-dependent Cox model

Another use case for the time-dependent Cox model is when the effect of the risk factors changes over time thus violating the proportional hazards assumption. In such case, a known function of time $g(t)$ (e.g., $\log(t)$) is added to the model[6]:

$$\lambda\left(t|\mathbf{Z}\right) = \lambda_0(t)\exp\left[\left(\boldsymbol{\beta}_1 + g(t)\boldsymbol{\beta}_2\right)^\top \mathbf{Z}\right]$$

The HR of the hazard function with a risk vector $\mathbf{Z}_1$ vs. that with a risk vector $\mathbf{Z}_2$ is:

$$\begin{aligned}\frac{\lambda\left(t|\mathbf{Z}_1\right)}{\lambda\left(t|\mathbf{Z}_2\right)} &= \frac{\lambda_0(t)\exp\left[\left(\boldsymbol{\beta}_1 + g(t)\boldsymbol{\beta}_2\right)^\top \mathbf{Z}_1\right]}{\lambda_0(t)\exp\left[\left(\boldsymbol{\beta}_1 + g(t)\boldsymbol{\beta}_2\right)^\top \mathbf{Z}_2\right]} \\ &= \exp\left[\boldsymbol{\beta}_1^\top\left(\mathbf{Z}_1 - \mathbf{Z}_2\right) + g(t)\boldsymbol{\beta}_2^\top\left(\mathbf{Z}_1 - \mathbf{Z}_2\right)\right]\end{aligned}$$

which depends on t if $\boldsymbol{\beta}_2$ is not zero.

[6] John P. Klein et al. *Survival Analysis - Techniques for Censored and Truncated Data*. P303

Time-dependent Cox model

Code example - Risk factor with two categories

```
proc phreg data = final;
  model intxsurv * dead(0) = trtgp ttrtgp / rl;
  ttrtgp = trtgp * log(intxsurv);
run;
```

Code example - Risk factor with more than two categories

```
proc phreg data = final;
  model intxsurv * dead(0) = trtgp2 ttrtgp2 trtgp3 ttrtgp3 / rl;
  if trtgp = 2 then trtgp2 = 1; else trtgp2 = 0;
  if trtgp = 3 then trtgp3 = 1; else trtgp3 = 0;
  ttrtgp2 = trtgp2 * log(intxsurv);
  ttrtgp3 = trtgp3 * log(intxsurv);
run;
```

Piecewise proportional hazards model

Even if the proportional hazards assumption does not hold for all time t, it is often possible to find it hold regionally:

$$\lambda\left(t|\mathbf{Z}\right) = \begin{cases} \lambda_0(t)\exp\left[\boldsymbol{\beta}_1^\top \mathbf{Z}\right] & t \leq \tau \\ \lambda_0(t)\exp\left[\boldsymbol{\beta}_2^\top \mathbf{Z}\right] & t > \tau \end{cases}$$

This is equivalent to:

$$\lambda\left(t|\mathbf{Z}\right) = \lambda_0(t)\exp\left[\boldsymbol{\beta}_1^\top \mathbf{Z}_1(t) + \boldsymbol{\beta}_2^\top \mathbf{Z}_2(t)\right]$$

where

$$\mathbf{Z}_1(t) = \begin{cases} \mathbf{Z} & t \leq \tau \\ \mathbf{0} & t > \tau \end{cases} \quad \mathbf{Z}_2(t) = \begin{cases} \mathbf{0} & t \leq \tau \\ \mathbf{Z} & t > \tau \end{cases}$$

The value of τ that yields the largest log partial likelihood is the optimal value of τ.

Piecewise proportional hazards model

Code example - Risk factor with two categories

```
proc phreg data = final;
  model intxsurv * dead(0) = trtgp_e trtgp_l / rl;
  tau = 6;
  if intxsurv <= tau then do;
    trtgp_e = trtgp;
    trtgp_l = 0;
  end;
  else do;
    trtgp_e = 0;
    trtgp_l = trtgp;
  end;
run;
```

How to check proportional hazards assumption?
What about a risk factor with more than two categories?

Stratified Cox model

$$\lambda_k(t|\mathbf{Z}) = \lambda_{0k}(t) \exp\left[\boldsymbol{\beta}^\top \mathbf{Z}\right], k = 1, \ldots, K$$

- Allows different baseline hazard functions $\lambda_{0k}(t)$ across strata.
- Factors do not satisfy the proportionality assumption can be stratified, and are no longer considered in the risk set $\mathbf{Z}$.
- Same coefficient $\boldsymbol{\beta}$ is assumed across strata for the remaining factors in $\mathbf{Z}$.
- Estimation of HRs for the stratified factors is not possible.

Code example

```
1 proc phreg data = final;
2   strata trtgp;
3   class sex race / ref = first param = ref;
4   model intxsurv * dead(0) = sex race / rl;
5 run;
```

Multivariate analysis using Cox model - Step by step

Step 1 - Check proportional hazards assumption

```
1 proc phreg data = final;
2   class trtgp sex race / ref = first param = ref;
3   model intxsurv * dead(0) = trtgp ttrtgp sex race;
4   ttrtgp = trtgp * log(intxsurv);
5 run;
```

Fit a piecewise model is the proportional hazards assumption is violated.

Step 2 - Model building

```
1 proc phreg data = final;
2   class trtgp sex race / ref = first param = ref;
3   model intxsurv * dead(0) = trtgp sex race / include = 1 selection =
      backward slstay = 0.05 detail;
4 run;
```

Multivariate analysis using Cox model - Step by step

Step 3 - Test interactions

```
proc phreg data = final;
  class trtgp race / ref = first param = ref;
  model intxsurv * dead(0) = trtgp black black * trtgp asian asian * trtgp;
  if race = 2 then black = 1; else black = 0;
  if race = 3 then asian = 1; else asian = 0;
run;
```

Step 4 - Final regression

```
proc phreg data = final;
  class trtgp race / ref = first param = ref;
  model intxsurv * dead(0) = trtgp black black * trtgp asian / rl;
  if race = 2 then black = 1; else black = 0;
  if race = 3 then asian = 1; else asian = 0;
run;
```

Estimation of baseline cumulative hazard function

The baseline cumulative hazard function $\Lambda_0(t) = \int_0^t \lambda_0(u)du$ for a Cox model can be estimated by the **Breslow estimator**[7]:

$$\hat{\Lambda}_0(t) = \sum_{t_j \leq t} \left[d_j \Big/ \sum_{T_i \geq t_j} \exp\left[\hat{\boldsymbol{\beta}}^\top \mathbf{Z}_i\right] \right]$$

t_j: jth distinct event time. d_j: Number of events at time t_j. T_i: Observed time for ith subject. $\mathbf{Z}_i$: Risk vector for ith subject.
When there is no covariate present, it reduces to the Nelson-Aalon estimator.

[7] John P. Klein et al. *Survival Analysis - Techniques for Censored and Truncated Data*. P283

Predicted survival function

The **predicted survival function** for a given risk vector $\mathbf{Z}$:

- Unstratified Cox model:

$$\begin{aligned}\hat{S}(t|\mathbf{Z}) &= \exp\left[-\int_0^t \hat{\lambda}(u|\mathbf{Z})du\right] \\ &= \exp\left[-\int_0^t \hat{\lambda}_0(u)\exp\left[\hat{\boldsymbol{\beta}}^\top \mathbf{Z}\right]du\right] \\ &= \exp\left[-\hat{\Lambda}_0(t)\exp\left[\hat{\boldsymbol{\beta}}^\top \mathbf{Z}\right]\right]\end{aligned}$$

  $\hat{\Lambda}_0(t)$ is the Breslow estimator.

- Stratified Cox model

$$\hat{S}_k(t|\mathbf{Z}) = \exp\left[-\hat{\Lambda}_{0k}(t)\exp\left[\hat{\boldsymbol{\beta}}^\top \mathbf{Z}\right]\right]$$

Direct adjusted survival function

The **direct adjusted survival function** averages the predicted survival functions of all subjects in the pooled sample. For a stratified Cox model:

$$\hat{\bar{S}}_k(t) = \frac{1}{n}\sum_{i=1}^{n} \hat{S}_k(t|\mathbf{Z}_i) = \frac{1}{n}\sum_{i=1}^{n} \exp\left[-\hat{\Lambda}_{0k}(t)\exp\left[\hat{\boldsymbol{\beta}}^\top \mathbf{Z}_i\right]\right]$$

where $\mathbf{Z}_i$ is the risk vector for ith subject in the pooled sample for $i = 1, \ldots, n$.

Code example

```
proc phreg data = final;
  strata trtgp;
  class sex race / ref = first param = ref;
  model intxsurv * dead(0) = sex race / rl;
  baseline out = outdata survival = _all_ diradj;
run;
```

Propensity score

- A **propensity score** is the probability of a subject being assigned to the treatment group $Z = 1$ given a set of risk factors $\mathbf{X}$:

$$e(\mathbf{X}) = P(Z = 1|\mathbf{X}) = E\left[Z|\mathbf{X}\right]$$

  where Z is the treatment assignment such that $1 =$ treated and $0 =$ untreated.
- Let Y_1 be the potential outcome for a subject if treated and Y_0 if untreated (regardless of the actual treatment assignment Z for the subject). If $(Y_1, Y_0) \perp\!\!\!\perp Z|\mathbf{X}$ and $0 < e(\mathbf{X}) < 1$, then $(Y_1, Y_0) \perp\!\!\!\perp Z|e(\mathbf{X})$.
- Allows estimation of **average treatment effect** (ATE) and **average treatment effect on the treated** (ATT) (also called weighting by odds):

$$\begin{aligned}\text{ATE} &= E\left[Y_1 - Y_0\right] \\ \text{ATT} &= E\left[Y_1 - Y_0|Z = 1\right]\end{aligned}$$

Propensity score matching

- Pairs of treated and untreated subjects are formed such that matched subjects have similar values of propensity scores.

  Greedy A treated subject is randomly selected. The untreated subject whose propensity score is closest to that of the selected treated subject is chosen for matching.

  Optimal Matches are formed to minimize the total within-pair difference of the propensity score.

- Allows estimation of ATT.
- Residual differences in baseline characteristics may still exist and can be benefited from using further regression adjustment on the risk factors.

Inverse probability of treatment weighting

Inverse probability of treatment weighting (IPTW) for ATE:

$$w_{j,\text{ATE}} = \frac{Z_j}{e_j} + \frac{1 - Z_j}{1 - e_j}$$

Stabilized IPTW for ATE

$$w^*_{j,\text{ATE}} = \frac{\pi Z_j}{e_j} + \frac{(1 - \pi)(1 - Z_j)}{1 - e_j}$$

ATT weighting:

$$w_{j,\text{ATT}} = Z_j + \frac{(1 - Z_j)e_j}{1 - e_j}$$

Z_j: Treatment assignment for jth subject. e_j: Propensity score for jth subject. π: Proportion of subjects in the treated group.

Propensity score - Steps

1. Logistic regression model building
   - Treatment group as dependent variable (0 vs. 1)
   - List of covariates to be considered
2. Logistic regression with selected covariates
   - Obtain predicted probability (propensity score) of being in treatment group 1 for all subjects
3. Use estimated propensity score in survival analysis model
   - Use propensity score for matching and treat matched pairs as clustered data in the Cox model
   - Use propensity score to calculate IPTW and use it as weight in the Cox model (similar to the use of survey sampling weight)
   - Use propensity score as a continuous factor directly in the Cox model

Propensity score - Step by step

Step 1 - Model building

```
1 proc logistic data = final;
2   class trtgp z1-z5 / ref = first param = ref;
3   model trtgp = z1-z5 / rl selection = stepwise slentry = 0.05 slstay = 0.05;
4 run;
```

Propensity score - Step by step

Step 2 - Obtaining propensity score

Matching

```
proc psmatch data = final region = cs;
  class trtgp z1-z5;
  psmodel trtgp(treated = '1') = z1-z5;
  match method = optimal(k = 1) exact = (z1-z2) stat = lps caliper = 0.25;
  assess lps var = (z1-z5) / plots = all weight = none;
  output out(obs = match) = outpsmatch matchid = matchid;
run;
```

Stabilized IPTW for ATE

```
proc psmatch data = final region = allobs(psmin = 0.05 psmax = 0.95);
  class trtgp z1-z5;
  psmodel trtgp(treated = '1') = z1-z5;
  assess lps var = (z1-z5) / plots = all weight = atewgt(stabilize = yes);
  output out(obs = region) = outiptw atewgt(stabilize = yes) = iptw;
run;
```

Propensity score - Step by step

Step 3 - Using the propensity score in survival analysis

Use matched pairs as clustered data

```
proc phreg data = outpsmatch covsandwich(agg);
  class trtgp z1-z10 / ref = first param = ref;
  model intxsurv * dead(0) = trtgp z1-z10 / rl;
  id matchid;
run;
```

Use IPTW as weight

```
proc surveyphreg data = outiptw varmethod = bootstrap;
  class trtgp z1-z10 / ref = first param = ref;
  model intxsurv * dead(0) = trtgp z1-z10 / rl;
  weight iptw;
run;
```

Right censoring vs. left truncation

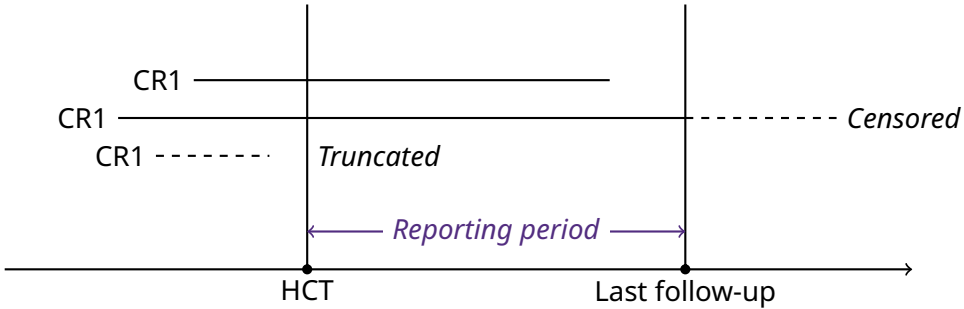
Right censoring

- The lifetime of a subject is observed up until certain (censoring) time.
- For those who are right-censored, their lifetimes are **partically** observable.

Left truncation

- A subject enters the study at a **delayed entry time** (left-truncation time).
- For those who are left-truncated, their lifetimes are **not** observable.

Left truncation - Example



The clock starts at CR1 while data is collected at HCT.

- Lifetimes of those who died between HCT and last follow-up are known.
- Lifetimes of those who were still alive at last follow-up are particially known (up till the last follow-up).
- Lifetimes of those who achieved CR1 with intention to recieve an HCT but did not eventually (e.g., early death) are not observable.

Adjustment for left-truncated data

When calculating number at risk Y at a given time t:

- Right censoring:

$$Y = \sum_{i=1}^{n} I(T_i \geq t)$$

  where T_i is the event time of ith subject in the sample for $i = 1, \ldots, n$.

- Right censoring and left truncation:

$$Y = \sum_{i=1}^{n} I(T_i \geq t > L_i)$$

  where T_i is the event time and L_i is the truncation/entry time of ith subject in the sample for $i = 1, \ldots, n$.

Left truncation - Example

L_i	T_i	δ_i
1	3	1
1	4	1
2	5	1
4	5	0
4.5	6	1
5.5	6	1
5.6	6	0
2	7	1
3.5	7	1
7.5	9	1
4.5	9	0

t	d	c	Y
0	0	0	0
3	1	0	4
4	1	0	4
5	1	1	6
6	2	1	6
7	2	0	3
9	1	1	2

What is the value of Y when $t = 5.5$?

Left truncation - Example (SAS)

Kaplan-Meier estimator

```
1 proc phreg data = final;
2   model indxsurv * dead(0) / entrytime = indxtx;
3   baseline out = outsurv survival = _all_ / method = pl;
4 run;
```

Cox proportional hazards model

```
1 proc phreg data = final;
2   class z1-z10 / ref = first param = ref;
3   model indxsurv * dead(0) = z1-z10 / rl entrytime = indxtx;
4 run;
```

`indxsurv` Time from diagnosis to death/last follow-up (event time)

`indxtx` Time from diagnosis to transplant (left-truncation time)

Power calculation and sample size

- **Power** Probability that the test correctly rejects the null hypothesis H_0 when the alternative hypothesis H_a is true.
- Three key components in power calculation:
  - Power (usually set at 80%)
  - Sample size
  - Expected difference in outcomes (based on published literature)
- Need two to calculate the third (usually sample size)

Power calculation - Example (SAS)

Code example

```
proc power;
  twosamplesurvival
  ntotal = 1861 groupweights = (19 1)
  /* Alternatively: groupns = 1768|93 */
  alpha = .05 sides = 2 test = logrank
  accrualtime = 0.01 followuptime = max
  curve('hct') = (1 to 6 by 1):(0.70 0.59 0.54 0.52 0.50 0.48)
  curve('chemo') = (1 to 3 by 1):(0.79 0.71 0.64)
  groupsurvival = 'hct'|'chemo'
  power = .;
run;
```