# Survival Analysis

P8108 Midterm - Chapters 1-4

Xuange Liang
xl3493@cumc.columbia.edu

## Ch1: Basic Definitions & Notation

**Notation**: $h(t) = \lambda(t)$ (hazard), $H(t) = \Lambda(t)$ (cumulative hazard)

**Key relationships**:
$S(t) = 1 - F(t) = e^{-H(t)}$; $F(t) = \int_0^t f(u)du$;
$f(t) = -S'(t)$
$h(t) = \frac{f(t)}{S(t)} = -\frac{d \log S(t)}{dt}$;
$H(t) = \int_0^t h(u)du = -\log S(t)$
$f(t) = h(t)S(t)$; $P(T > s + t | T > s) = \frac{S(s+t)}{S(s)}$

**Median**: $t_{0.5}$ where $S(t_{0.5}) = 0.5$; **Mean**: $E(T) = \int_0^\infty S(t)dt$

**Properties**: $S(0) = 1$, $S(\infty) = 0$, non-increasing, right-continuous. $h(t) \geq 0$ (can be $> 1$)

## Ch1: Types of Censoring

**Right censoring** (most common):
Event occurs after observation period ends
Observe $X = \min(T, C)$ where $C$=censoring time
$\delta = I(T \leq C)$: event indicator (1=event, 0=censored)

**Left censoring**: event before study starts
**Interval censoring**: event in known interval

**Type I censoring**: fixed censoring time
**Type II censoring**: study ends after fixed # events

**Random censoring**: censoring time random

**Left truncation**: only observe subjects surviving past entry $L$. Risk set: only those entered and not yet failed/censored. Differs from left censoring

## Ch1: Censoring Assumptions

**Non-informative (independent) censoring**:
Censoring time independent of failure time
$P(T > t | T > u, \text{censored at } u) = P(T > t | T > u)$
REQUIRED for valid KM estimator

**Informative censoring**:
Censoring related to failure mechanism
Violates standard methods - need special approaches

**Example of informative**: patients withdraw due to worsening condition

**Censoring distribution**: Reverse KM: treat censoring as "event". Check if censoring differs between groups. $\hat{G}(t) = P(C > t)$

## Ch2: Kaplan-Meier Estimator

**Product-limit estimator**:

$$\hat{S}(t) = \prod_{t_i \leq t} \left(1 - \frac{d_i}{r_i}\right) = \prod_{t_i \leq t} \frac{r_i - d_i}{r_i}$$

where:
$t_1 < t_2 < \cdots$: ordered distinct event times
$d_i$: # events at $t_i$
$r_i$: # at risk at $t_i$ (alive just before $t_i$)

**Risk set at $t_i$**: Include those who fail at $t_i$ and censored after $t_i$; exclude those censored at $t_i$ or failed before $t_i$

**Key**: $\hat{S}(t)$ step function, decreases only at event times

**Greenwood's variance formula**:

$$\text{Var}(\hat{S}(t)) = \hat{S}(t)^2 \sum_{t_i \leq t} \frac{d_i}{r_i(r_i - d_i)}$$

**Standard error**:

$$SE(\hat{S}(t)) = \hat{S}(t) \sqrt{\sum_{t_i \leq t} \frac{d_i}{r_i(r_i - d_i)}}$$

**KM-based cumulative hazard**:
$\hat{\Lambda}_{KM}(t) = -\log \hat{S}_{KM}(t)$. Alternative to NA.
Inverse: $\hat{S}_{KM}(t) = e^{-\hat{\Lambda}_{KM}(t)}$

**Variance of $\log \hat{S}(t)$**:

$$\text{Var}[\log \hat{S}(t)] = \sum_{t_i \leq t} \frac{d_i}{r_i(r_i - d_i)}$$

Note: NO $\hat{S}^2$ factor! Used in log transformation CI

**CI for $\hat{S}(t)$ - Three methods**:

**1. Plain (linear)**: $\hat{S}(t) \pm 1.96 SE(\hat{S}(t))$
Problem: may exceed $[0, 1]$

**2. Log transformation**: $SE[\log \hat{S}(t)] = \frac{SE(\hat{S})}{\hat{S}(t)}$
CI: $\exp[\log \hat{S}(t) \pm 1.96 SE(\log \hat{S})]$

**3. Log-log (preferred, default)**:
$\theta = \log[-\log \hat{S}(t)]$
$SE(\theta) = \frac{\sqrt{\sum d_i/(r_i(r_i - d_i))}}{|\log \hat{S}(t)|}$
CI: $\exp\{-\exp(\theta \pm 1.96 SE(\theta))\}$
Always stays in $[0, 1]$, best for extreme $\hat{S}(t)$ values

**Percentiles**: $p$-th: smallest $t_p$ where $\hat{S}(t_p) \leq 1 - p$. Median: $\hat{S}(t_{0.50}) = 0.50$. If $\hat{S}(t)$ never reaches threshold, undefined

**Backward calculation**: Given $\hat{S}(t)$ product form & $n$, find $d_i$, $r_i$, $c_i$:

**Step-by-step**: 1) From $\hat{S}(t_i)/\hat{S}(t_{i-1}) = (r_i - d_i)/r_i$, read denominator $\Rightarrow r_i$, then $d_i = r_i -$ numerator. 2) Compute $c_i$: if $r_1 < n$, then $n - r_1$ censored before $\tau_1$; otherwise $c_i = (r_i - d_i) - r_{i+1}$. 3) Verify: $\sum d_i + \sum c_i = n$ (critical!)

**Ex (Sample Q1)**: $n = 13$, $\hat{S}(\tau_1) = 10/11 \Rightarrow r_1 = 11, d_1 = 1, c_{\text{early}} = 13 - 11 = 2$.
$\hat{S}(\tau_2) = (10/11)(9/10) \Rightarrow r_2 = 10, d_2 = 1, c_1 = (11 - 1) - 10 = 0$. Continue. Final check: $\sum d_i = 6, \sum c_i = 7, 6 + 7 = 13$ ✓

**Trap**: If product starts $< 1$, early censoring! $r_1 =$ first denominator, NOT $n$

**Backward calc**: $n = 10$, $\hat{S}(\tau_1) = \frac{8}{9}$, $\hat{S}(\tau_2) = \frac{8}{9} \times \frac{6}{7}$, $\hat{S}(\tau_3) = \frac{8}{9} \times \frac{6}{7} \times \frac{4}{5} \Rightarrow$
$r_1 = 9, d_1 = 1, c_0 = 1$; $r_2 = 7, d_2 = 1, c_1 = 1$;
$r_3 = 5, d_3 = 1, c_2 = 1$; $c_3 = 4$. Check: $3 + 7 = 10$ ✓

**RMST**: $\text{RMST}(\tau) = \int_0^\tau \hat{S}(t)dt$ (area under KM to $\tau$). Use when median not reached

**KM example**: Data: 2, 3+, 5, 6. $t = 2$: $r = 4, d = 1, \hat{S}(2) = \frac{3}{4} = 0.75$. $t = 3$: censored, $\hat{S} = 0.75$. $t = 5$:
$r = 2, d = 1, \hat{S}(5) = 0.75 \times \frac{1}{2} = 0.375$. Greenwood: $\text{Var}[\hat{S}(5)] = (0.375)^2[\frac{1}{12} + \frac{1}{2}] = 0.082$. Check: $\sum d_i + \sum c_i = n$

## Ch2: Other Estimators

**Nelson-Aalen (NA) cumulative hazard**:

$$\hat{H}_{NA}(t) = \sum_{t_i \leq t} \frac{d_i}{r_i}$$

Variance: $\text{Var}[\hat{H}_{NA}(t)] = \sum_{t_i \leq t} \frac{d_i}{r_i^2}$

**Fleming-Harrington estimator**:

$$\hat{S}_{FH}(t) = \exp(-\hat{H}_{NA}(t))$$

Close to $\hat{S}_{KM}$ when hazards small. Uses NA for cumulative hazard

**Comparison**: NA: direct $H(t)$ estimate; KM: direct $S(t)$ estimate. $\hat{\Lambda}_{KM}(t) \approx \hat{H}_{NA}(t)$ for small hazards

**Life table (actuarial)**: For grouped data in $[t_j, t_{j+1})$: effective risk set $r_j^* = r_j - c_j/2$, then $\hat{q}_j = d_j/r_j^*$, $\hat{S}(t_{j+1}) = \hat{S}(t_j)(1 - \hat{q}_j)$

**NA vs KM relationship**:
KM: $\hat{S}(t) = \prod(1 - \frac{d_i}{r_i})$
NA: $\hat{H}_{NA}(t) = \sum \frac{d_i}{r_i}$, $\hat{S}_{FH}(t) = e^{-\hat{H}_{NA}(t)}$

When $d_i/r_i$ small: $1 - \frac{d_i}{r_i} \approx e^{-d_i/r_i}$
$\Rightarrow \hat{S}_{KM}(t) \approx \hat{S}_{FH}(t)$ and $-\log \hat{S}_{KM}(t) \approx \hat{H}_{NA}(t)$

## Ch3: Comparing Two Groups

**Logrank test (Mantel-Haenszel)**:
$H_0: S_1(t) = S_2(t)$ for all $t$
At each event time $t_i$:

$$E_{1i} = \frac{r_{1i} \cdot d_i}{r_i} \quad \text{(expected in group 1)}$$

$$V_i = \frac{r_{1i} r_{2i} d_i (r_i - d_i)}{r_i^2 (r_i - 1)} \quad \text{(variance)}$$

where $r_i = r_{1i} + r_{2i}$, $d_i = d_{1i} + d_{2i}$
Test statistic:

$$\chi^2_{LR} = \frac{[\sum_i (O_{1i} - E_{1i})]^2}{\sum_i V_i} \sim \chi^2_1$$

**Weights**: Logrank=equal; Wilcoxon:
$\chi^2_W = [\sum r_i (O_{1i} - E_{1i})]^2 / \sum r_i^2 V_i$ (early); $G^{\rho,\gamma}$:
$w_i = \hat{S}^\rho (1 - \hat{S})^\gamma$, $(\rho, \gamma)$: $(0, 0)$=logrank, $(1, 0)$=early, $(0, 1)$=late

**Test selection**: Logrank when PH holds. Curves cross: Wilcoxon or weighted tests. With confounders: stratified logrank

**Stratified logrank**:
For $K$ strata, sum over strata:

$$\chi^2 = \frac{[\sum_{k=1}^K \sum_i (O_{1ik} - E_{1ik})]^2}{\sum_{k=1}^K \sum_i V_{ik}}$$

**Logrank steps**: 1) Pool all unique event times. 2) At each $t_i$: count $r_{1i}, r_{2i}, d_{1i}, d_{2i}$. 3) Calculate $E_{1i}, V_i$. 4) Sum: $O_1 = \sum d_{1i}$, $E_1 = \sum E_{1i}$, $V = \sum V_i$. 5) $\chi^2 = (O_1 - E_1)^2/V \sim \chi^2_1$

**Key points**: Pool times first. Skip censoring times. $r_i$ from pooled data. Sum over ALL event times

## Ch4: Cox Proportional Hazards

**Model form**:

$$h(t, Z) = h_0(t) \exp(\beta_1 Z_1 + \cdots + \beta_p Z_p)$$
$$h(t, Z) = h_0(t) e^{\beta^T Z}$$

**Components**:
$h_0(t)$: baseline hazard (when all $Z = 0$)
$Z = (Z_1, \ldots, Z_p)^T$: covariate vector
$\beta = (\beta_1, \ldots, \beta_p)^T$: coefficients

**Semi-parametric**:
No parametric form assumed for $h_0(t)$
Estimate $\beta$ without specifying $h_0(t)$

**Proportional hazards property**:

$$\frac{h(t, Z^{(1)})}{h(t, Z^{(0)})} = \frac{h_0(t) e^{\beta^T Z^{(1)}}}{h_0(t) e^{\beta^T Z^{(0)}}} = e^{\beta^T (Z^{(1)} - Z^{(0)})}$$

Ratio independent of $t$ (constant over time)

**Hazard ratio (HR)**:

$$HR = e^{\beta^T (Z^{(1)} - Z^{(0)})}$$

For binary $Z$: $HR = e^\beta$
For continuous $Z$ ($c$-unit change): $HR = e^{\beta c}$

**Interpretation**:
$\beta > 0$: $HR > 1$, increased hazard (worse survival)
$\beta < 0$: $HR < 1$, decreased hazard (better survival)
$\beta = 0$: $HR = 1$, no effect

**HR meaning**: HR=2 means hazard doubles; HR=0.5 means hazard halves. NOT survival probability!

**Log-linear model**:
$\log(HR) = \beta^T Z$
Linear in log-hazard scale

**Partial likelihood**:
Cox (1972, 1975) developed partial likelihood
Conditions on risk sets at each failure time
Avoids specifying $h_0(t)$

$$L(\beta) = \prod_{i=1}^{D} \frac{\exp(\beta^T Z_{(i)})}{\sum_{j \in R(t_{(i)})} \exp(\beta^T Z_j)}$$

where $R(t_i)$ = risk set at time $t_i$

**Estimation**:
Maximize $\log L(\beta)$ using Newton-Raphson
Obtain $\hat{\beta}$ and $\text{Var}(\hat{\beta})$ from information matrix

## Ch4: Inference for Cox Model

**Three tests for** $H_0: \beta = 0$: 1. **Wald**: $\chi^2_W = \frac{\hat{\beta}^2}{Var(\hat{\beta})} \sim \chi^2_1$ 2. **Score**: $\chi^2_S = U(0)^T I(0)^{-1} U(0)$ (logrank for single binary covariate) 3. **Partial LR**: $\chi^2_{LR} = 2[\log L(\hat{\beta}) - \log L(\beta_0)]$ (most reliable)
Generally: $\chi^2_W \le \chi^2_{LR} \le \chi^2_S$ for small effects

**CI for HR**: 95% CI $= e^{\hat{\beta} \pm 1.96 SE(\hat{\beta})}$. CI excludes $1 \Leftrightarrow p < 0.05$

**Overall test**: For $H_0: \beta_1 = \cdots = \beta_p = 0$, use partial LR with $p$ df

**Confounding**: $> 10\%$ change in $\hat{\beta}$ suggests confounding

**Nested model comparison**:
For nested models (Model 0 $\subset$ Model 1):
$H_0$: extra parameters $= 0$
$\chi^2_{LR} = 2[\log L_1(\hat{\beta}_1) - \log L_0(\hat{\beta}_0)]$
df = # extra parameters in Model 1

**Individual coefficient test**:
Wald: $Z = \frac{\hat{\beta}_j}{SE(\hat{\beta}_j)} \sim N(0,1)$ or $\chi^2 = Z^2 \sim \chi^2_1$
95% CI: $\hat{\beta}_j \pm 1.96 \times SE(\hat{\beta}_j)$
For HR: $e^{\hat{\beta}_j \pm 1.96 \times SE(\hat{\beta}_j)}$

## Ch4: Tied Times & Other Topics

**Handling ties**: Breslow (fast), Efron (better, default in R), Exact (slow). Use Efron for most cases

**Breslow method**: At time $t_i$ with $d_i$ tied deaths:

$$L_i(\beta) = \frac{\prod_{j \in D_i} \exp(\beta^T Z_j)}{\left[\sum_{k \in R_i} \exp(\beta^T Z_k)\right]^{d_i}}$$

Approximation: uses same risk set for all tied deaths

**Efron method**: Better approximation for ties
Let $S_R = \sum_{k \in R_i} e^{\beta^T Z_k}$, $S_D = \sum_{j \in D_i} e^{\beta^T Z_j}$

$$L_i(\beta) = \frac{\prod_{j \in D_i} e^{\beta^T Z_j}}{\prod_{l=1}^{d_i} \left[S_R - \frac{l-1}{d_i} S_D\right]}$$

Adjusts risk set by removing fraction of tied deaths

**Practical notes**: Breslow/Efron give similar results when few ties. Efron preferred when >5% ties

## Ch4: Stratified Cox Model

When PH assumption violated for a covariate:

**Stratified model**:

$$h_k(t, Z) = h_{0k}(t) e^{\beta^T Z}$$

Different baseline $h_{0k}(t)$ for stratum $k$
Common $\beta$ across strata

**Stratification vs adjustment**:
Stratify: don't estimate effect, allow different baselines
Adjust: estimate effect, assume common baseline

**Interaction**:
$h(t|Z_1, Z_2) = h_0(t) e^{\beta_1 Z_1 + \beta_2 Z_2 + \beta_3 Z_1 Z_2}$. Effect of $Z_1$ at $Z_2 = c$: $HR = e^{\beta_1 + \beta_3 c}$. Test: $H_0: \beta_3 = 0$

**Multiple covariates**:
$HR = \exp[\beta_1 (Z_1^{(1)} - Z_1^{(0)}) + \beta_2 (Z_2^{(1)} - Z_2^{(0)})]$

**Categorical (>2 levels)**: Use dummy variables. 3-level: 2 indicators. Overall test: LR with 2 df

**Variance-Covariance matrix**:
For $\hat{\beta} = (\hat{\beta}_1, \hat{\beta}_2)^T$:

$$\text{Var}(\hat{\beta}_1 + \hat{\beta}_2) = \text{Var}(\hat{\beta}_1) + \text{Var}(\hat{\beta}_2) + 2\text{Cov}(\hat{\beta}_1, \hat{\beta}_2)$$

Important for linear combinations of parameters (e.g., interaction effects)

**Baseline estimation**: $\hat{h}_0(t_i) = \frac{d_i}{\sum_{j \in R_i} e^{\hat{\beta}^T Z_j}}$;
$\hat{H}_0(t) = \sum \hat{h}_0(t_i)$; $\hat{S}_0(t) = e^{-\hat{H}_0(t)}$

**Predicted survival**: $\hat{S}(t|Z) = [\hat{S}_0(t)]^{\exp(\hat{\beta}^T Z)}$.
Precision: # events, not subjects

## Ch4: Examples & Interpretation

**Basic HR**: $h(t) = h_0(t) e^{0.5 X_1 + 0.03 X_2}$ ($X_1$=trt, $X_2$=age). Trt vs control: $HR = e^{0.5} = 1.65$; 10y age ↑: $HR = e^{0.3} = 1.35$; Trt 60yo vs control 50yo: $HR = e^{0.5+0.3} = e^{0.8} = 2.23$

**CI to SE**: HR=1.5 (1.2, 1.9) $\Rightarrow$
$\hat{\beta} = \log(1.5) = 0.405$;
$SE(\hat{\beta}) = [\log(1.9) - \log(1.2)]/[2(1.96)] = 0.117$;
$Z = 0.405/0.117 = 3.46$, $p < 0.001$

**Stratified**: $h(t) = h_{0,stage}(t) e^{-0.971 X_E + 0.003 X_A}$. ER+ vs ER- (same age/stage): $HR = e^{-0.971} = 0.38$. Different stages: CANNOT compare!

**Interaction**: $h(t) = h_0(t) e^{\beta_1 Z_1 + \beta_2 Z_2 + \beta_3 Z_1 Z_2}$. Effect of $Z_1$ at $Z_2 = 0$: $HR = e^{\beta_1}$; at $Z_2 = 1$: $HR = e^{\beta_1 + \beta_3}$

**Reading KM plots**: Step function; median = first $t$ where $\hat{S}(t) \le 0.5$; crossing curves $\Rightarrow$ PH violated; use formal tests (logrank/Cox); CI overlap is not a significance test;
# at risk shown in table below

## Ch4: Checking PH Assumption

**PH assumption**: $h_1(t)/h_0(t) = HR$ (constant) $\Rightarrow S_1(t) = [S_0(t)]^{HR}$ (multiplicative, NOT additive)

**On KM plot** ($S(t)$ vs $t$): Curves NOT parallel! Distance changes over time. Non-crossing $\Rightarrow$ PH plausible. Crossing $\Rightarrow$ PH violated

**On log-log plot** ($\log[-\log S(t)]$ vs $t$): Curves ARE parallel if PH holds. Vertical distance = $\log(HR)$. Why:
$\log[-\log S_1(t)] = \log[-\log S_0(t)] + \log(HR)$

**Tests**: 1) KM non-crossing. 2) Log-log parallel. 3) `cox.zph()` in R

**If violated**: Stratify (different $h_0$, don't estimate effect)

## Key Formulas & Values

**Critical values**: $z_{0.975} = 1.96$, $z_{0.995} = 2.576$, $\chi^2_{1,0.95} = 3.84$, $\chi^2_{2,0.95} = 5.99$

**Useful logs/exp**: $\log(2) = 0.693$, $\log(0.5) = -0.693$, $e^{0.5} \approx 1.65$, $e^{-0.5} \approx 0.61$, $e^1 \approx 2.72$, $e^{-1} \approx 0.37$

**More useful values**: $\log(3) = 1.099$, $\log(0.38) \approx -0.97$, $e^{0.3} \approx 1.35$, $e^{-0.97} \approx 0.38$

**Conversions**: $\beta = \log(HR)$; $HR = e^{\beta}$;
$\hat{H}(t) = -\log \hat{S}(t)$; $\hat{S}(t) = e^{-\hat{H}(t)}$

**Percentage change in hazard**:
HR = 1.5 means 50% increase: $(HR - 1) \times 100\%$
HR = 0.6 means 40% decrease: $(1 - HR) \times 100\%$

**Cox prediction**: Individual survival $\hat{S}(t|Z) = [\hat{S}_0(t)]^{\exp(\hat{\beta}^T Z)}$. For $c$-unit change: $HR = e^{\beta c}$. Multiple changes:
$HR = \exp[\beta_1 \Delta Z_1 + \beta_2 \Delta Z_2]$

**Variance formulas**: Greenwood uses $\hat{S}^2$. For $\log \hat{S}$: $\text{Var}[\log \hat{S}] = \sum_{t_i \le t} \frac{d_i}{r_i(r_i - d_i)}$ (no $\hat{S}^2$ factor!)

**Partial likelihood (KEY CONCEPT)**: At each failure time $t_i$, conditional prob that individual $i$ fails given one failure in risk set $R_i$:

$$L_i(\beta) = \frac{\exp(\beta^T Z_i)}{\sum_{j \in R_i} \exp(\beta^T Z_j)}$$

Full partial likelihood: $L(\beta) = \prod_{i=1}^{D} L_i(\beta)$ over all $D$ events. Maximize $\log L(\beta)$ to get $\hat{\beta}$

**Log partial likelihood**:

$$\log L(\beta) = \sum_{i=1}^{D} \left[\beta^T Z_i - \log\left(\sum_{j \in R_i} \exp(\beta^T Z_j)\right)\right]$$

Score: $U(\beta) = \frac{\partial \log L}{\partial \beta}$; Information: $I(\beta) = -\frac{\partial^2 \log L}{\partial \beta \partial \beta^T}$

**Baseline hazard (Breslow estimator)**: After getting $\hat{\beta}$:

$$\hat{h}_0(t_i) = \frac{d_i}{\sum_{j \in R_i} \exp(\hat{\beta}^T Z_j)}$$

Then $\hat{H}_0(t) = \sum_{t_i \le t} \hat{h}_0(t_i)$, $\hat{S}_0(t) = \exp(-\hat{H}_0(t))$

**Relationship to KM**: When $\beta = 0$ (no covariates), Cox model reduces to KM estimator

**Exponential**: $h(t) = \lambda$ (constant), $S(t) = e^{-\lambda t}$, $H(t) = \lambda t$. Median= $0.693/\lambda$. Mean= $1/\lambda$. Memoryless: $P(T > s + t | T > s) = P(T > t)$. MLE: $\hat{\lambda} = d / \sum_{i=1}^{n} X_i$

**Weibull**: $h(t) = \alpha \lambda t^{\alpha - 1}$, $S(t) = e^{-\lambda t^{\alpha}}$. $\alpha > 1$: ↑ hazard; $\alpha < 1$: ↓ hazard; $\alpha = 1$: exponential. Check: plot $\log H(t)$ vs $\log t$ linear

**Log-normal**: Non-monotone hazard (increases then decreases)

## Interpretation Templates

**Compare treatments**:
**No diff**: "No difference between [A] vs [B] in [outcome]. KM curves [nearly identical], HR=[val] (95% CI: [...]), $p = [val]$ (not sig)."
**Sig diff**: "[Advantage/Disadvantage] to [A] vs [B] in [outcome]. KM curves [separate], HR=[val] (95% CI: [...]), $p = [val]$ (sig)."
Note: HR<1 favors exposed (lower hazard); HR>1 favors unexposed

**PH check**:
**OK**: "KM curves [parallel/non-crossing], PH reasonable. [Further diagnostics possible.]"
**Violated**: "KM curves [cross/converge], PH may be violated. [Consider stratified/time-dep.]"

**Cox model format**:
Binary $W$: $\lambda(t, W) = \lambda_0(t) e^{\beta W}$, $\beta = \log(HR)$
Continuous $X$: HR for $c$-unit ↑: $e^{\beta c}$
Multiple: $\lambda(t, Z) = \lambda_0(t) e^{\beta_1 Z_1 + \beta_2 Z_2 + \cdots}$
Interaction: Effect of $Z_1$ at $Z_2 = 0$: $HR = e^{\beta_1}$; at $Z_2 = 1$: $HR = e^{\beta_1 + \beta_3}$

**HR interpretation**: HR>1: $[(HR - 1) \times 100]\%$ ↑ hazard; HR<1: $[(1 - HR) \times 100]\%$ ↓ hazard

## Quick Reference

**Test equivalences**: Score test for single binary covariate = Logrank test; Logrank = $G^{0,0}$ (equal weights); Wilcoxon = $G^{1,0}$ (early differences)

**Key inequalities**: $\chi^2_W \le \chi^2_{LR} \le \chi^2_S$; For small $x$: $1 - x \approx e^{-x}$, $\log(1 + x) \approx x$; Delta method: $\text{Var}[g(X)] \approx [g'(\mu)]^2 \text{Var}(X)$

**Assumptions**: KM: non-informative censoring; Logrank: PH + non-informative censoring; Cox PH: proportional hazards + non-informative censoring