SESSION 4:
INTRODUCTION TO COX REGRESSION

Module 12: Introduction to Survival Analysis
Summer Institute in Statistics for Clinical Research
University of Washington
July, 2017

Barbara McKnight, Ph.D.
Professor
Department of Biostatistics
University of Washington

OVERVIEW

- Session 1
  - Introductory examples
  - The survival function
  - Survival Distributions
  - Mean and Median survival time
- Session 2
  - Censored data
  - Risk sets
  - Censoring Assumptions
  - Kaplan-Meier Estimator and CI
  - Median and CI
- Session 3
  - Two-group comparisons: logrank test
  - Trend and heterogeneity tests for more than two groups
- Session 4
  - Introduction to Cox regression

OUTLINE

- Motivation:
  - Confounding in observational studies
  - Stratified randomization designs
- Cox Regression model
  - Coefficient interpretation
  - Estimation and testing
  - Relationship to 2- and K-sample tests
  - Examining non-proportionality
- Examples throughout

OUTLINE

- **Motivation:**
  - **Confounding in observational studies**
  - **Stratified randomization designs**
- Cox Regression model
  - Coefficient interpretation
  - Estimation and testing
  - Relationship to 2- and K-sample tests
  - Examining non-proportionality
- Examples throughout

CONFOUNDING

- Observational data: sometimes observed associations between an explanatory variable and outcome can be due to their joint association with another variable.
  - Age related to both sex and risk of death.
  - Other examples?

PRECISION IN RCTS

- Because of randomization, confounding/imbalance usually not an issue except in small trials.
- As in linear regression, regression models for censored survival data allow group comparisons among subjects with similar values of adjustment or "precision" variables (more later).
- Fairer and possibly more powerful comparison as long as adjustment variables are not the result of treatment.

STRATIFIED RANDOMIZATION

- For strong predictors: concern about possible randomization imbalance
  - Clinic or center
  - Stage of disease
  - Sex
  - Age
- Adjust for stratification variables in analysis
  - More powerful if predictors are strong
  - Same conditioning as the sampling

OUTLINE

- Motivation:
  - Confounding in observational studies
  - Stratified randomization designs
- **Cox Regression model**
  - **Coefficient interpretation**
  - **Estimation and testing**
  - **Relationship to 2- and K-sample tests**
  - **Examining non-proportionality**
- Examples throughout

COX REGRESSION MODEL

- Usually written in terms of the hazard function
- As a function of independent variables $x_1, x_2, \ldots x_k$,

$$\lambda(t) = \lambda_0(t)e^{\beta_1 x_1 + \cdots + \beta_k x_k}$$

↑ relative risk / hazard ratio

$$\log \lambda(t) = \log \lambda_0(t) + \beta_1 x_1 + \cdots + \beta_k x_k$$

↑ intercept

RELATIVE RISK / HAZARD RATIO

$$\lambda(t|x_1, \ldots, x_k) = \lambda_0(t)e^{\beta_1 x_1 + \cdots + \beta_k x_k}$$

$$\frac{\lambda(t|x_1,\ldots,x_k)}{\lambda(t|0,\ldots,0)} = e^{\beta_1 x_1 + \cdots + \beta_k x_k}$$

REGRESSION MODELS

LS Linear Regression: $Y = \beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k + \epsilon$

Linear:	$Y \sim N(\mu, \sigma^2)$	$\mu = EY = \beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k$
Cox:	$T \sim S(t)$	$\lambda(t) = \lambda_0(t) e^{\beta_1 x_1 + \cdots + \beta_k x_k}$
	↑	↑
	Distribution of outcome variable	Dependence of distribution on $x_1, \ldots x_k$

PROPORTIONAL HAZARDS MODEL

$$\lambda(t) = \lambda_0(t) e^{\beta_1 x_1 + \cdots + \beta_k x_k}$$

Interpretation of e^{β_1} in general:

"Relative risk (or hazard ratio) associated with a one unit higher value of x_1, holding $x_2, \ldots, x_k$ constant".

$\lambda(t)$ for $x_1 + 1$:	$\lambda_0(t) e^{\beta_1 (x_1 + 1) + \cdots + \beta_k x_k}$
$\lambda(t)$ for x_1:	$\lambda_0(t) e^{\beta_1 x_1 + \cdots + \beta_k x_k}$
ratio:	$e^{\beta_1 (x_1 + 1 - x_1)} = e^{\beta_1}$

EXAMPLE

Single binary x:

$$x = \begin{cases} 1 & \text{Test treatment} \\ 0 & \text{Standard treatment} \end{cases}$$

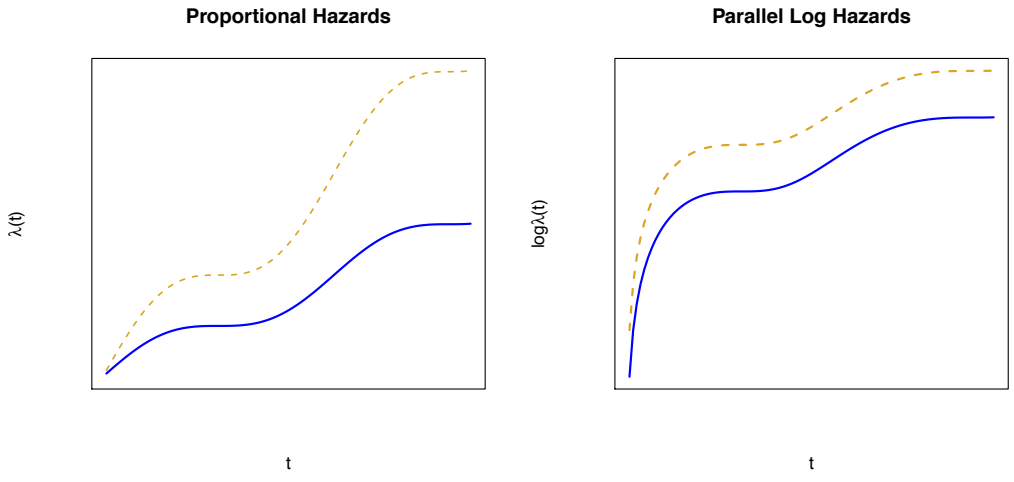
$$\lambda(t) = \lambda_0(t)e^{\beta x}$$

Interpretation of e^{β}:

"Relative risk (or hazard ratio) comparing test treatment to standard".

$$\begin{aligned} \lambda(t) \text{ for } x = 1: \quad & \lambda_0(t)e^{\beta \cdot 1} = \lambda_0(t)e^{\beta} \\ \lambda(t) \text{ for } x = 0: \quad & \lambda_0(t)e^{\beta \cdot 0} = \lambda_0(t) \\ \text{ratio:} \quad & e^{\beta(1-0)} = e^{\beta} \end{aligned}$$

EXAMPLE



RELATIONSHIP TO SURVIVAL FUNCTION

Single binary x:

$$x = \begin{cases} 1 & \text{Test treatment} \\ 0 & \text{Standard treatment} \end{cases}$$

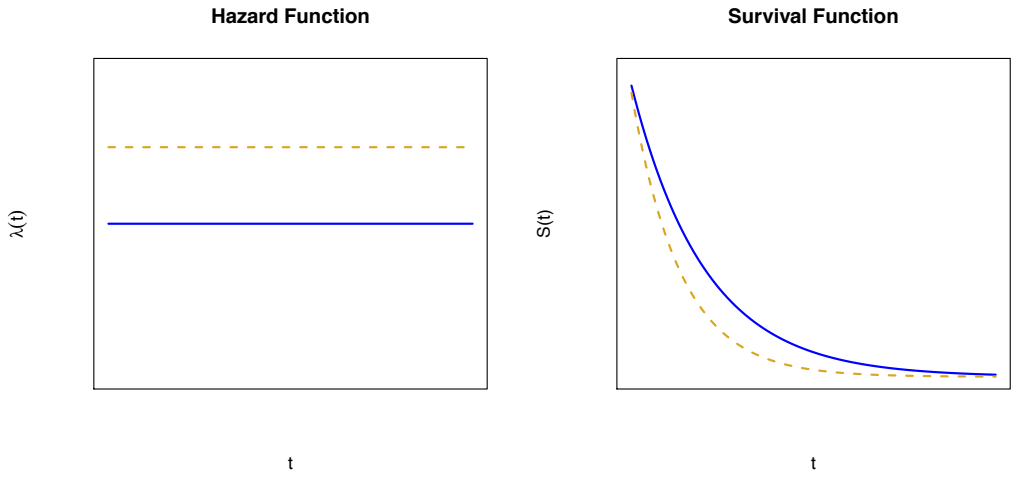
$$\lambda(t) = \lambda_0(t)e^{\beta x} \implies S(t) = [S_0(t)]^{e^{\beta x}}$$

In terms of $S_0(t)$:

$S(t)$ for $x = 1$: $\quad [S_0(t)]^{e^{\beta \cdot 1}} = [S_0(t)]^{e^{\beta}}$

$S(t)$ for $x = 0$: $\quad [S_0(t)]^{e^{\beta \cdot 0}} = [S_0(t)]^{1} = S_0(t)$

PICTURE



ESTIMATES AND CONFIDENCE INTERVALS

- We estimate β by maximizing the "partial likelihood function"
- Requires iteration on computer
- $\hat{\beta}$ is a MPLE (Maximum Partial Likelihood Estimator)
- We do not need to estimate $\lambda_0(t)$ to do this

- Most packages will estimate se($\hat{\beta}$) using the information matrix from this PL.
- 95% CI for β: $(\hat{\beta} - 1.96\text{se}(\hat{\beta}),\ \hat{\beta} + 1.96\text{se}(\hat{\beta}))$
- 95% CI for RR = e^{β} : $(e^{\hat{\beta}-1.96\text{se}(\hat{\beta})}, e^{\hat{\beta}+1.96\text{se}(\hat{\beta})})$

PARTIAL LIKELIHOOD

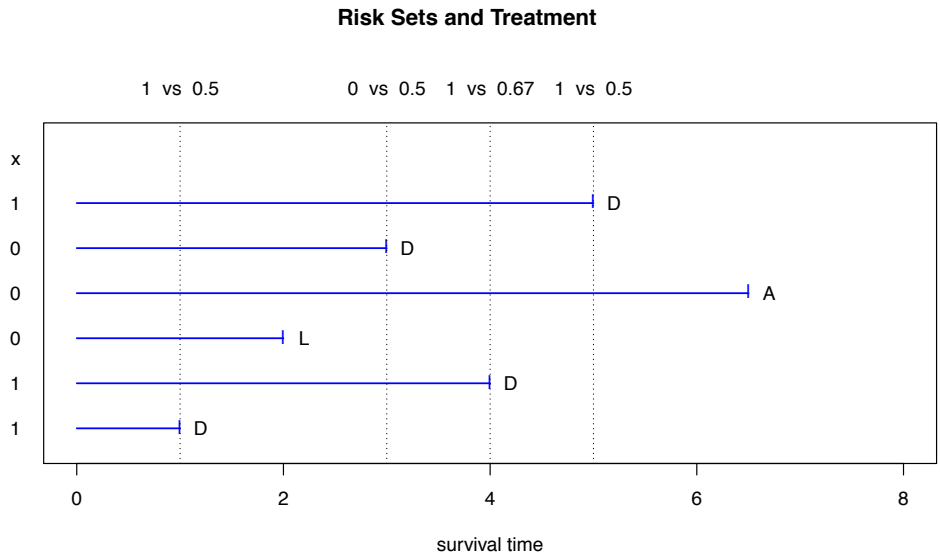
Data for the i^{th} subject: $(t_i, \delta_i, x_{1i}, \ldots x_{ki})$

For subject with the j^{th} ordered failure time : $(t_{(j)}, 1, x_{1(j)}, \ldots, x_{k(j)})$

$$\text{PL}(\beta_1, \ldots, \beta_k) = \prod_{j=1}^{J} \frac{e^{\beta_1 x_{1(j)} + \cdots + \beta_k x_{k(j)}}}{\sum_{i:t_i \geq t_{(j)}} e^{\beta_1 x_{1i} + \cdots + \beta_k x_{ki}}}$$

- $(\hat{\beta}_1, \ldots, \hat{\beta}_k)$ are the values of $(\beta_1, \ldots, \beta_k)$ that maximize PL$(\beta_1, \ldots, \beta_k)$. (MPLEs)
- Compares x values for the subject who failed at time $t_{(j)}$ to those of all subjects at risk at time $t_{(j)}$.
- Does not depend on the values of the t_i, only on their order.
- Does not depend on $\lambda_0(t)$.

RISK SET PICTURE



FULL LIKELIHOOD

$$
\begin{aligned}
L(\beta, \lambda_0(t)) &= \prod_{\text{Failures}} \Pr[T = t_i] \prod_{\text{Censorings}} \Pr[T > t_i] \\
&= \prod_{\text{Failures}} \lambda(t_i|x_i) S(t_i|x_i) \prod_{\text{Censorings}} S(t_i|x_i) \\
&= \prod_{i=1}^{n} [\lambda(t_i|x_i)]^{\delta_i} S(t_i|x_i) \\
&= \prod_{i=1}^{n} [\lambda_0(t_i) e^{\beta x_i}]^{\delta_i} e^{-\int_0^{t_i} \lambda_0(s) e^{\beta x} ds}
\end{aligned}
$$

PARTIAL LIKELIHOOD

Let H_t represent the entire history of failure, censoring and x in the sample before time t.

Then the likelihood can be rewritten as follows:

$$
\begin{aligned}
L(\beta, \lambda_0(t)) &= \prod_{j=1}^{J} \Pr[i^{th} \text{ subject fails at } t_{(j)} | H_{t_{(j)}}, \text{ some subject fails at } t_{(j)}] \cdot \\
&\qquad \Pr[H_{t_{(j)}}, \text{ some subject fails at } t_{(j)}] \\
&= \prod_{j=1}^{J} \frac{\lambda(t_{(j)}|x_{(j)})}{\sum_{i:t_i \geq t_{(j)}} \lambda(t_{(j)}|x_i)} \cdot \prod_{j=1}^{J} \Pr[H_{t_{(j)}}, \text{ some subject fails at } t_{(j)}] \\
&= \prod_{j=1}^{J} \frac{\lambda_0(t_{(j)}) e^{\beta x_{(j)}}}{\sum_{i:t_i \geq t_{(j)}} \lambda_0(t_{(j)}) e^{\beta x_i}} \cdot \prod_{j=1}^{J} \Pr[H_{t_{(j)}}, \text{ some subject fails at } t_{(j)}] \\
&= \prod_{j=1}^{J} \frac{e^{\beta x_{(j)}}}{\sum_{i:t_i \geq t_{(j)}} e^{\beta x_i}} \cdot \prod_{j=1}^{J} \Pr[H_{t_{(j)}}, \text{ some subject fails at } t_{(j)}] \\
&= \underbrace{\qquad\qquad}_{\text{Partial Likelihood}} \quad \underbrace{\qquad\qquad\qquad\qquad\qquad\qquad}_{\text{Depends on } \lambda_0(\cdot) \text{ and } \beta}
\end{aligned}
$$

Depends only on β

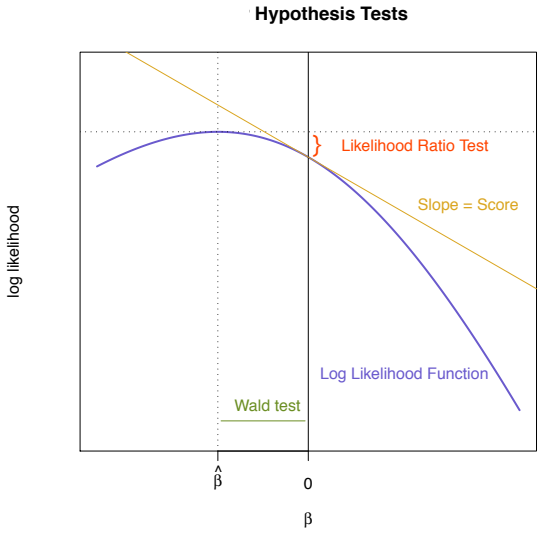
HYPOTHESIS TESTS

Three tests of $H_0 : \beta = 0$ are possible:

1. Wald test: $\dfrac{\hat{\beta}}{\text{se}(\hat{\beta})}$
2. (Partial) Likelihood ratio test
3. Score test: ($\approx$ logrank test)

Likelihood ratio test is best, but requires fitting full ($\beta = \hat{\beta}$) and reduced ($\beta = 0$) models.

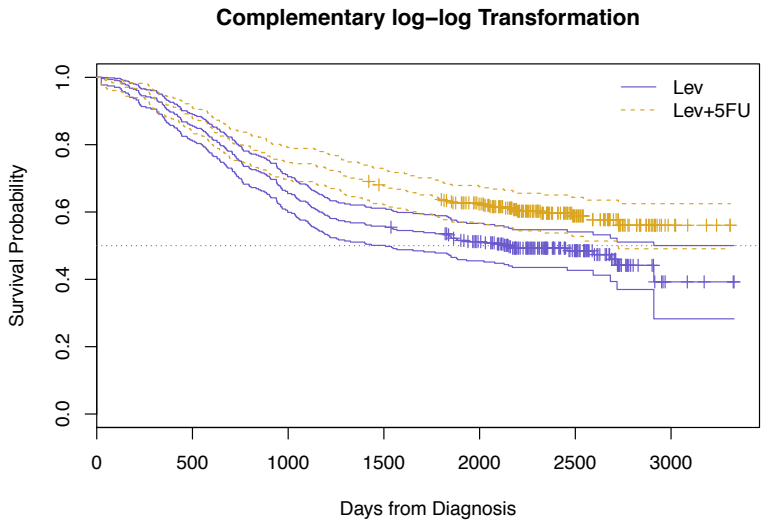
LIKELIHOODS AND TESTS



COLON CANCER EXAMPLE

- Clinical trial at Mayo Clinic
- Stage B_2 and C colon cancer patients; adjuvant therapy
- Three arms
  - Observation only
  - Levamisole (stage C only)
  - 5-FU + Levamisole at Mayo Clinic
- Stage C patients only
- Two treatment arms only

COLON CANCER EXAMPLE



COLON CANCER EXAMPLE

Variable	n	Deaths	Hazard ratio	CI	P-value
Levamisole Only	310	161	1.0 (reference)	--	--
Levamisole + 5FU	304	123	0.71	(0.56, 0.90)	.004

Q: Which group has better survival?

A:

TEST COMPARISON

Test	Statistic	P-value
Wald's	8.13	.004
Score	8.21	.004
Likelihood Ratio	8.21	.004

Two-sided tests

ANOTHER EXAMPLE

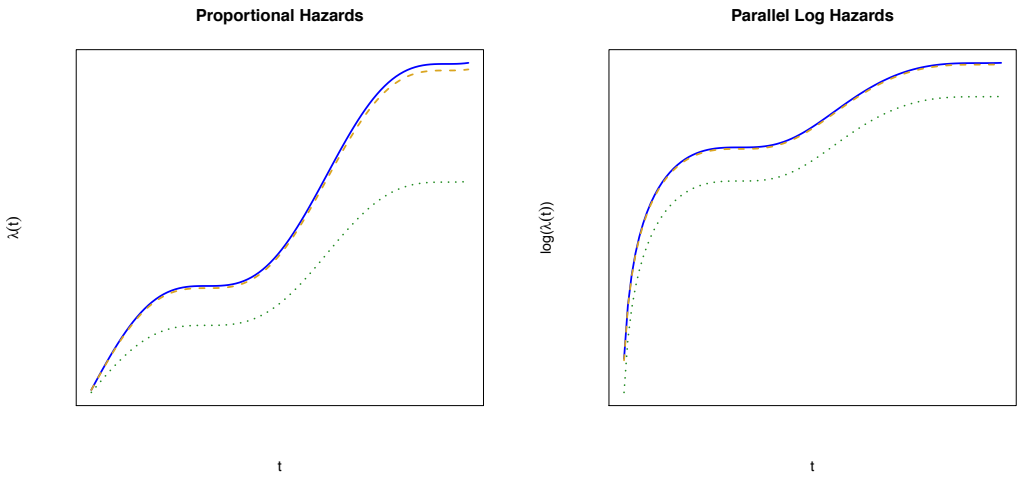
Three groups: use indicators for two

$$x_1 = \begin{cases} 1 & \text{Levamisole Only} \\ 0 & \text{otherwise} \end{cases} \qquad x_2 = \begin{cases} 1 & \text{Levamisole + 5FU} \\ 0 & \text{otherwise} \end{cases}$$

Model: $\lambda(t) = \lambda_0(t)e^{\beta_1 x_1 + \beta_2 x_2}$

RRs:	Levamisole Only	vs.	Observation	e^{β_1}
	Levamisole + 5FU	vs.	Observation	e^{β_2}
	Levamisole + 5FU	vs.	Levamisole Only	$e^{\beta_2 - \beta_1}$

HEURISTIC HAZARDS



COLON CANCER

Variable	n	Deaths	Hazard Ratio	95% CI	P-value
Observation Only	315	168	1.0 (reference)	--	--
Levamisole Only	310	161	0.97	(0.78, 1.21)	0.81
Levamisole + 5FU	204	123	0.69	(0.55, 0.87)	0.002

Q: Which group has best survival?

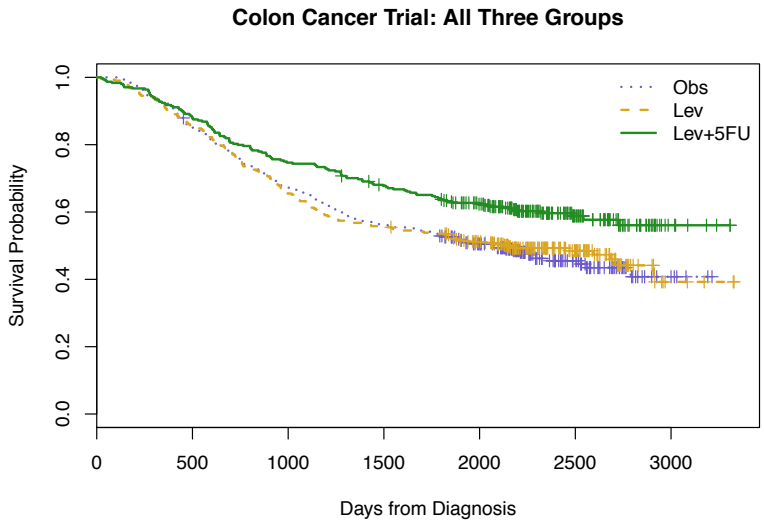
A:

TEST COMPARISON

Test	Statistic	P-value
Wald's	11.56	.003
Score	11.68	.003
Likelihood Ratio	12.15	.002

Same hypothesis as 3-group heterogeneity test. Score test is same in large samples.

COLON CANCER TRIAL DATA



TREND

- When there are several groups, it is sometimes of interest to test whether risk increases from one group to the next:
  - Several dose groups
  - Other ordered variable
  - Example: tumor differentiation
- For $x = \begin{cases} 1 & \text{well differentiated} \\ 2 & \text{moderately differentiated} \\ 3 & \text{poorly differentiated} \end{cases}$

$$\text{Model: } \lambda(t) = \lambda_0(t)e^{\beta x}$$

- Score test is the same as the trend test
- Could use other values for x (actual dose levels)

TREND

For $x = \begin{cases} 1 & \text{well differentiated} \\ 2 & \text{moderately differentiated} \\ 3 & \text{poorly differentiated} \end{cases}$

$$\text{Model: } \lambda(t) = \lambda_0(t)e^{\beta x}$$

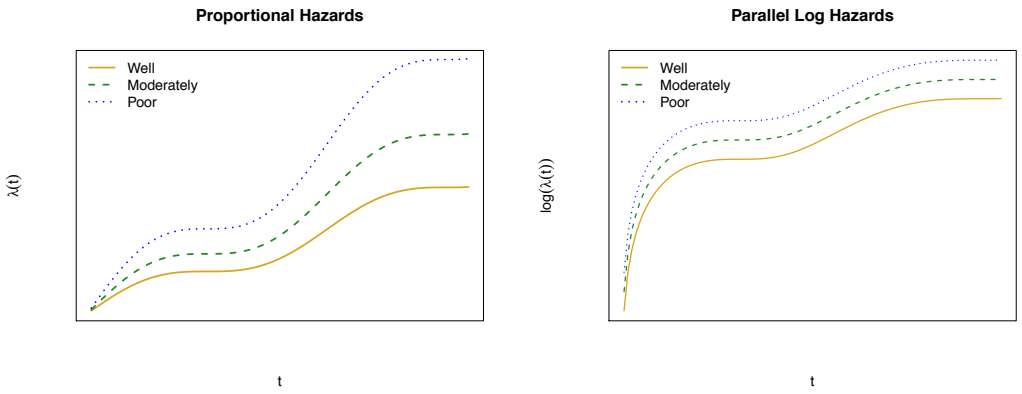
Interpretation of e^{β}: HR associated with the comparison of one worse differentiation group to one better:

- poorly differentiated to moderately differentiated, or
- moderately differentiated to well differentiated

Q: What is HR comparing poorly differentiated to well differentiated?

A:

TREND



TREND WITH DIFFERENTIATION

One presentation based entirely on trend ("grouped linear") model:

	Hazard Ratio	95% CI
One category worse differentiation (well, moderately, poor)	1.4	(1.1, 1.8)
P = .003 (trend)		

I prefer presenting hazard ratios and CI's based on dummy variable model, and providing P-value for trend.

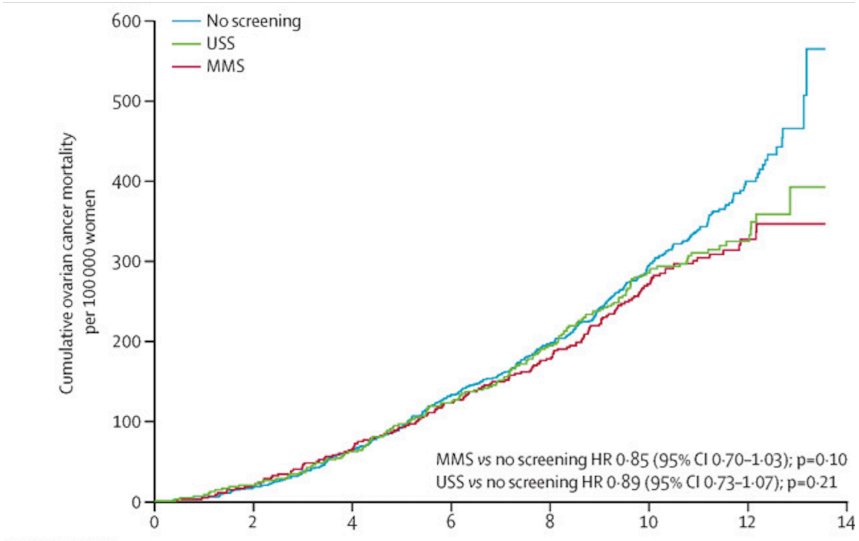
TREND WITH DIFFERENTIATION

My preferred presentation based on dummy variable mode with trend P-value:

	n	Deaths	Hazard Ratio	95% CI
Well differentiated	66	26	1.0 (reference)	--
Moderately differentiated	434	196	1.2	(0.80, 1.8)
Poorly differentiated	98	54	1.8	(1.2, 3.0)
P = .003 (trend)				

I usually would not present this for an *a priori* trend hypothesis, but for comparison here, the heterogeneity P-value (2 df) is 0.009.

OVARIAN CANCER SCREENING TRIAL



PROPORTIONAL HAZARDS

- One way to examine evidence against proportional hazards is to look at plots of scaled Schoenfeld residuals and perform tests based on them.
- For each failing subject there is a Schoenfeld residual for each x variable in the model.
- At the subject's failure time, the residual measures how the value of x for the subject who fails differs from a weighted average of x values for those still at risk. (Weights depend on estimated HR for each subject at risk).
- If consistently high or low over an interval of time, this is evidence that the hazard at that time is even higher (lower) for the subject with that x than the model indicates.

SCHOENFELD RESIDUALS

Formula for Schoenfeld residuals

Let $r_i(t) = e^{\hat{\beta}x_i(t)}$ be the estimated hazard ratio for the i^{th} subject at t compared to $x(t) = 0$.

Then for $\overline{x}(\hat{\beta}, t) = \frac{\sum_{\text{at risk at } t} r_i(t)x_i(t)}{\sum_{\text{at risk at } t} r_i(t)}$,

The Schoenfeld residual for the k^{th} subject failing at time t is given by $x_k(t) - \overline{x}(\hat{\beta}, t)$.

The scaled Schoenfeld residual is the Schoenfeld residual divided by a variance estimate.

SCHOENFELD RESIDUALS

- Grambsch and Therneau (1994) showed that the scaled Schoenfeld residual measures the deviation of a time-dependent log hazard ratio $\beta(t)$ from time-constant $\hat{\beta}$.
- Can use linear regression comparing scaled Schoenfeld residuals to functions of time to examine evidence for lack of constant hazard ratio over time.
- Grambsch PM, Therneau TM. Biometrika. 1994 Sep 1;81(3):515–526.

COLON CANCER TRIAL DATA

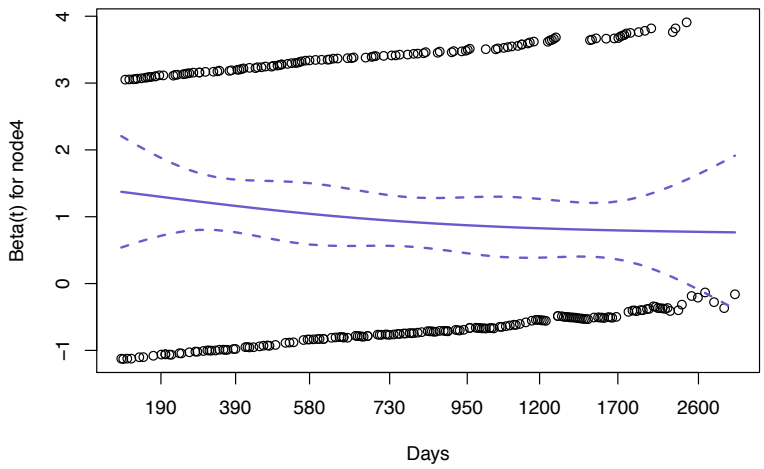
Observation Arm Omitted

	$\hat{\beta}$	exp($\hat{\beta}$)	se($\hat{\beta}$)	z	Pr(>|z|)
5FU + Lev	-0.34	0.71	0.12	-2.83	0.0064
4+ Nodes Pos	0.98	2.67	0.12	8.08	<0.0001

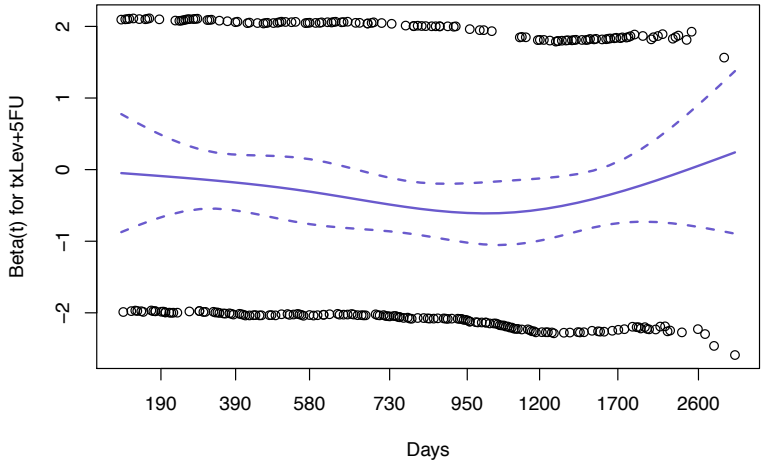
$e^{\beta_{Rx}}$ CI: (0.5629, 0.9008)

LRT: 8.098 on 1 df, P = 0.0044

FOR NODE 4 POSITIVITY



FOR TREATMENT



TEST FOR NON-PROPORTIONALITY

Variable	P-value
node4	0.158
txLev+5FU	0.560

No strong evidence for non-proportionality based on scaled Schoenfeld residuals correlation with "time" S(t).

TO WATCH OUT FOR:

- Coefficients in Cox regression are positively associated with risk, not survival.
  - Positive β means large values of x are associated with shorter survival.
- Without certain types of time-dependent covariates, Cox regression does not depend on the actual times, just their order.
  - Can add a constant to all times to remove zeros (some packages remove observations with time = 0) without changing inference
- For LRT, nested models must be compared based on same subjects.
  - If some values of variables in larger model are missing, these subjects must be removed from fit of smaller model.
- Hazards may not always be proportional