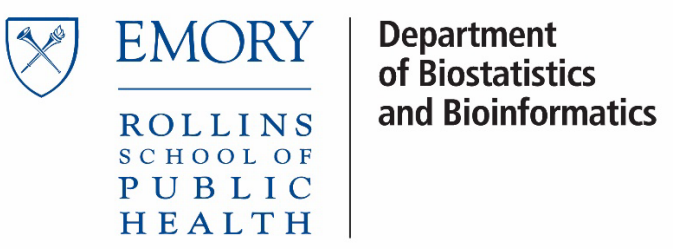


BIOS 522: Survival Analysis Methods

Reading 7:

Cox model diagnostics

This week, we will study the various types of residuals available for examining the Cox regression model. We will discuss different methods for assessing the validity of the proportional hazards assumption. We will learn strategies for model building and selection.

Part 1. Introduction to diagnostics

Goals of diagnostics

There are two basic assumptions for the Cox model. The first assumption is that the *functional form of the covariates is correctly specified*. For example, do continuous covariates need to be transformed? Do we need to include interaction terms? This assumption is standard in all regression models.

The second basic assumption is that of *proportional hazards*, meaning that the effect of covariates on the hazard is constant over time. This assumption is unique to the Cox proportional hazards model.

We can use both diagnostic plots and hypothesis testing-based procedures to examine our assumptions. We also use these diagnostic procedures to detect **outliers** and **influential observations**.

Simple setting

In the simplest Cox models where there are only a few combinations of categorical covariates with a relatively large subsample within each of these groups, we can check the overall fit of the model by comparing the survival (or cumulative hazard) functions predicted by the Cox model with the Kaplan-

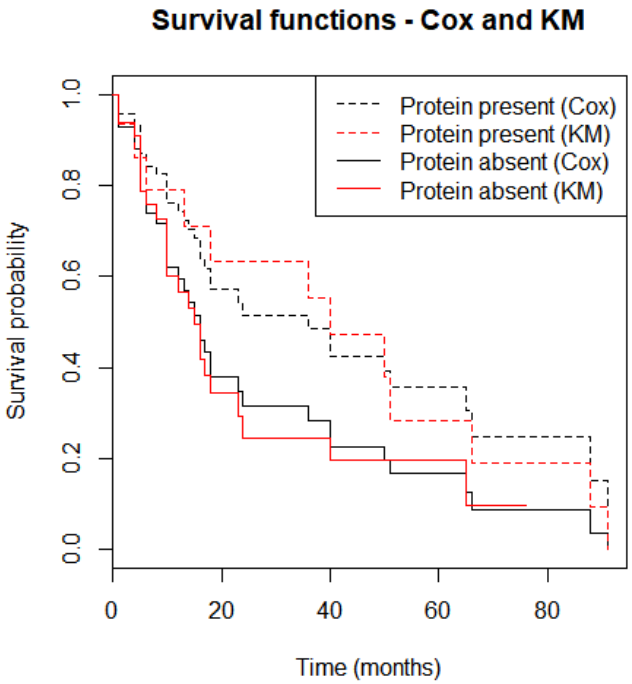
Meier (or Nelson-Aalen) curves for each subgroup. We view the curves side by side or overlay them on a single plot.

Example: Consider data on the survival of 48 patients with multiple myeloma – a malignant disease characterized by the accumulation of abnormal white blood cells in the bone marrow. The primary endpoint was the time in months from diagnosis until death from multiple myeloma.

At the time of diagnosis, data were collected on patient age, sex (1=male, 2=female), the levels of blood urea nitrogen (Bun), serum calcium (Ca), and hemoglobin (Hb), the percentage of plasma cells in the bone marrow (Pcells) and an indicator variable (Protein) that denotes whether or not the Bence-Jones protein was present in the urine (0=absent, 1=present).

The main purpose of the analysis was to investigate of the effect of the risk factors (Bun, Ca, Hb, Pcells, and protein) on survival time. The effects of these risk factors may be modified by age and sex.

Consider a Cox model with a single predictor for the presence of the Bence-Jones protein. To assess whether the data conform to a Cox model, we can overlay plots of Cox model predicted survival with the Kaplan-Meier curves fitted to each of the two groups.



We see relatively good agreement between the Cox and Kaplan-Meier curves, although there is some divergence in the protein present group between 20 and 40 months. This supports the use of the Cox model for this analysis.

This approach of comparing the Cox predicted survival with the Kaplan-Meier curve in each group is not very useful in more complex models where there are many combinations of covariates in the study population with only a small sample size at each combination. Where there is a small sample for each combination, the Kaplan-Meier curve will be unstable. This approach is also not as useful for continuous covariates, which would need to broken out into groups. A more flexible approach to goodness-of-fit is based on residuals.

Part 2. Residuals

Review and overview

Residuals typically measure the difference between the model predicted value and the observed value. For linear regression, the residual r_i is calculated as:

$$r_i = Y_i - \hat{Y}_i$$

where Y_i is the observed value of the outcome and $\hat{Y}_i$ is the predicted value for participant i.

Because of censoring, residuals for survival data are somewhat different than for other types of models. If T_i^* is a censoring time ($\delta_i = 0$), calculating a residual as the difference between the observed time T_i^* and a model-predicted time (e.g. median survival) is not meaningful. T_i^* is the duration of follow-up and not a direct measure of survival, particularly if T_i^* is a censoring time.

There are several types of residuals for a Cox model that are designed to handle censoring in the data. We will discuss the following methods:

- Cox-Snell (generalized) residuals
- Martingale residuals
- Deviance residuals
- Schoenfeld residuals

Cox-Snell (generalized) Residuals

The generalized residuals devised by Cox and Snell rely on the following result. If survival time T_i has cumulative hazard function $H_i(t)$, then the transformed random variable $H_i(T_i)$ follows an exponential distribution with rate $\lambda = 1$, i.e.:

$$H_i(T_i) \sim Exp(\lambda = 1)$$

Given a fitted Cox model with estimated cumulative hazard $\widetilde{H}(t)$ (e.g. estimated by the Breslow estimator), we can assess whether the fitted model reflects the true underlying model by calculating $\widetilde{H}_i(T_i)$ and comparing these to an $Exponential(\lambda = 1)$.

The **generalized** or **Cox-Snell residual** r_{Ci} is defined as follows:

$$r_{Ci} = \widetilde{H}_i(T_i^*)$$

Because the cumulative hazard function for an $Exp(\lambda = 1)$ is $H(t) = t$, we can plot the cumulative hazard function of the Cox-Snell residuals and compare it to a straight line with slope 1 that goes through the origin.

Example: Consider a Cox model with a single continuous covariate measuring blood urea nitrogen. The estimated log hazard ratio is $\hat{\beta}_{Bun} = 0.02$, and the estimated hazard ratio is $\exp(\hat{\beta}_{Bun}) = 1.02$. For each 1 mg/dL increase in blood urea nitrogen, the hazard of death due to multiple myeloma increases by a multiplicative factor of 1.02.

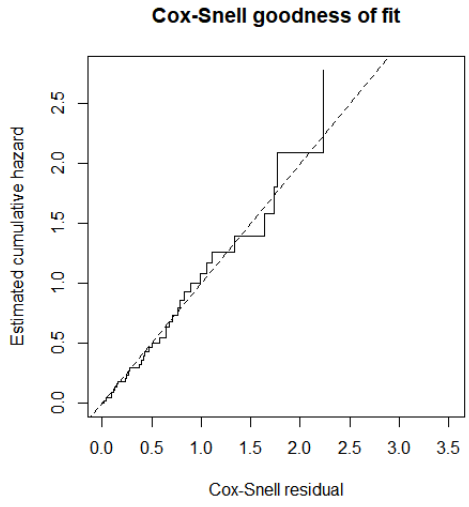
Patient 1 has blood urea nitrogen of 25 mg/dL. This patient failed at 13 months. Based on the fitted Cox model, their predicted survival at 13 months is 0.648. Their predicted cumulative hazard at 13 months is $-\log(0.648) = 0.434$.

Patient 2 has blood urea nitrogen of 13 mg/dL. This patient was censored at 52 months. Their predicted survival at 52 months is 0.326. Their predicted cumulative hazard at 52 months is $-\log(0.326) = 1.121$.

These two results are summarized in the following table:

Patient (i)	Time (T_i^*)	Status (δ_i)	BUN ($X_{i,Bun}$)	$\hat{S}_i(T_i^*)$	$r_{Ci} = \widehat{H}_i(T_i^*)$
1	13	1	25	0.648	0.434
2	52	0	13	0.326	1.121

We repeat this process for everyone in the dataset. We can generate a cumulative hazard function for our Cox-Snell residuals. A straight line with slope 1 that passes through the origin is added for reference.



We see that the residuals have a cumulative hazard function that roughly follows that of an $Exponential(\lambda = 1)$ distribution, suggesting good model fit.

Cox-Snell residuals are used primarily to assess the overall goodness of fit. It should be noted that it will take a particularly ill-fitting model for the Cox-Snell residuals to deviate significantly from a straight line. It is also not uncommon to see jumps occurring at the tail of the graph, due to variability in estimating the model parameters.

Martingale residuals

One criticism of Cox-Snell residuals is that they do not adequately distinguish between observed failure times and censoring times. One can modify the Cox-Snell residuals to explicitly account for censoring. We can also center them so that they have a mean of zero.

Martingale residuals r_{Mi} are defined for the ith individual as:

$$r_{Mi} = \delta_i - r_{Ci}$$

where δ_i is the event indicator. Note that the Cox-Snell residual r_{Ci} can also be interpreted as the expected number of events at the failure/censoring time based on the fitted model. Thus, the Martingale residual r_{Mi} can be viewed as the difference between the observed number of events (δ_i =0 or 1) for subject i between time 0 and T_i^*, and the expected number based on the fitted model.

Example: Consider our multiple myeloma dataset with Cox model for blood urea nitrogen. The Martingale residuals are calculated in the right-hand column.

Patient (i)	Time (T_i^*)	Status (δ_i)	BUN ($X_{i,Bun}$)	$r_{Ci} = \hat{H}_i(T_i^*)$	$r_{Mi} = \delta_i - r_{Ci}$
1	13	1	25	0.434	0.566
2	52	0	13	1.121	-1.121

■

Martingale residuals can take values between $-\infty$ and 1, and have mean 0.

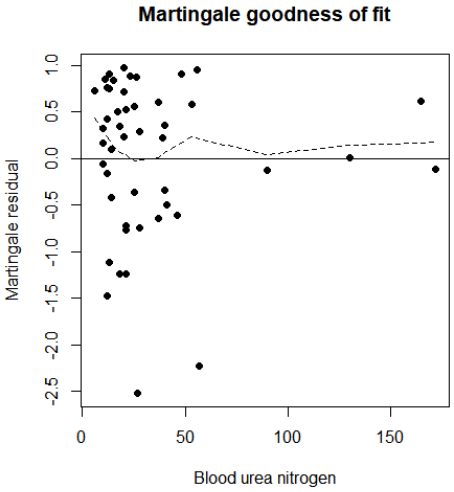
Martingale residuals can be used to identify outliers. Large negative residuals indicate individuals with long survival times who were predicted, based on covariates, to fail early on. Residuals at or near 1 indicate individuals with short survival times who were predicted, based on covariates, to survive a long time.

Martingale residuals can also be used to assess the functional form of a particular covariate (e.g., is a linear term adequate?). We can plot the Martingale residuals against our covariate X_j to check whether there is any unexplained relationship between the covariate and survival. If the linear model is adequate, the residuals should be randomly scattered.

Another strategy is to fit the model with all other covariates except X_j. We can examine the scatter plot of the Martingale residuals against X_j, adding a

smoothed curve (e.g. LOWESS) if helpful. This allows us to visualize the functional form of X_j. If the curve is already linear, then no transformation is needed.

Example: Consider our multiple myeloma dataset with Cox model for blood urea nitrogen. We can plot the Martingale residuals against the covariate $X_{i,Bun}$. A smoothing curve (dashed line) is overlaid on the plot.



The Martingale residuals shown do not indicate any significant deviation from random scatter. If we did see a shape in the plot, like a U-shape, then we should reconsider our model for blood urea nitrogen. Assuming that the effect is linear on the log hazard ratio scale may not be appropriate. We could consider fitting a polynomial or spline.

■

For models with many covariates $X_{i1}, \ldots, X_{ik}$, we can continue to examine each covariate one at a time. We may also wish to assess the overall model fit in a single plot. We can calculate the **linear predictor** X_{Pi}, which is the combined effect of all covariates:

$$X_{Pi} = \hat{\beta}_1 X_{i1} + \cdots \hat{\beta}_k X_{ik}$$

We can generate a plot with the linear predictor X_{Pi} on the x-axis and the Martingale residual on the y-axis. Again, we are looking for deviation from random scatter.

Deviance residuals

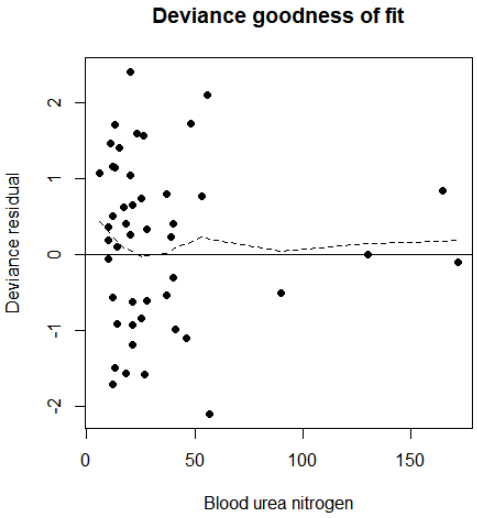
One problem with the Martingale residuals is that they tend to be asymmetric around 0, with larger negative values than positive values. One solution is to use deviance residuals.

Deviance residuals r_{Di} are defined for the ith individual as:

$$r_{Di} = sign(r_{Mi})\sqrt{-2[r_{Mi} + \delta_i \log(\delta_i - r_{Mi})]}$$

where r_{Mi} is the Martingale residual, and the *sign()* function takes a value of +1 for positive Martingale residuals and -1 for negative Martingale residuals.

Because they are designed to be symmetric around 0, deviance residuals are easier to visually examine than Martingale residuals when looking for outliers. They can be applied in the same way as Martingale residuals.



Example: Deviance residuals are plotted for the multiple myeloma model with blood urea nitrogen. Again, these do not show any deviation from random scatter.

■

Schoenfeld residuals

Cox-Snell, Martingale, and deviance residuals yield one residual per individual, including censored individuals. Schoenfeld residuals follow a different structure. They are:

- Only defined for failures.
- They are not defined for censored observations.
- Instead of there being a single Schoenfeld residual per individual, each individual has *multiple* Schoenfeld residuals – one for *each covariate*.

For a model with only a single covariate, the **Schoenfeld residual** r_{Si} for person i who failed at time T_i is:

$$r_{Si} = X_i - \frac{\sum_{i'=1}^{n_t} X_{i'} \exp(\hat{\beta} X_{i'})}{\sum_{i'=1}^{n_t} \exp(\hat{\beta} X_{i'})}$$

The second term is a weighted average of the covariate values X_i' for the n_t individuals at risk at T_i, with weights determined by each person's relative hazard $\exp(\hat{\beta} X_{i'})$. In essence, for each failure time, we calculate the covariate value for the individual who *failed* (observed value) minus its expected value given the model and the risk set at the time.

When there is more than one covariate in the model, the Schoenfeld residual will be a vector with one component for each variable in the model. For a model with k covariates, the Schoenfeld residual r_{Sij} for covariate j for person i who failed at time T_i is:

$$r_{Sij} = X_{ij} - \frac{\sum_{i'=1}^{n_t} X_{i'j} \exp(\hat{\beta} X_{i'j})}{\sum_{i'=1}^{n_t} \exp(\hat{\beta} X_{i'j})}$$

Example: Continuing with our previous multiple myeloma model with blood urea nitrogen, individual $i = 7$ in the dataset fails at time $T_i^* = 66$ months. At that time, there are $n_t = 4$ individuals remaining at risk. Their blood urea nitrogen values are summarized below. Given the fitted coefficient $\hat{\beta} = 0.0202$

Patient (i)	Time (T_i^*)	Status (δ_i)	BUN ($X_{i,Bun}$)	exp ($\hat{\beta} X_{i,Bun}$)
7	66	1	21	1.527
19	76	0	12	1.274
21	88	1	21	1.527
36	91	1	27	1.723

The Schoenfeld residual for person $i = 7$ is:

$$r_{Si} = 21 - \frac{\big(21(1.527) + 12(1.274) + 21(1.527) + 27(1.723)\big)}{(1.527 + 1.274 + 1.527 + 1.723)}$$
$$= 21 - 20.8 = 0.2$$

■

A further modification is to divide the Schoenfeld residual by its estimated standard error. This is known as the **scaled Schoenfeld residual**. One advantage of the scaled residuals is that they look more normally distributed, especially when the covariate is binary.

Schoenfeld residuals are particularly useful for examining the proportional hazards assumption. Note that we plotted the Martingale and deviance residuals against the *covariate* or *linear predictor*. Schoenfeld residuals are commonly plotted against *time*. In principle, the Schoenfeld residuals are independent of time. A plot that shows a non-random pattern against time is evidence of violation of the proportional hazard assumption.

Part 3. Examining the proportional hazards assumption

Overview

The proportional hazards assumption is central to Cox regression. The assumption is that the ratio of hazard functions in the populations being compared is constant in time. Covariates impact the survival time through a multiplicative effect on the hazard function.

There are many methods available for evaluating whether the proportional hazards assumption is appropriate for a particular model and dataset.

We will discuss the following methods:

- Simple graphical checks of Kaplan-Meier curves
- Log-log survival curves
- Plotting Schoenfeld residuals

- Testing Schoenfeld residuals slope
- Time-by-covariate interactions

Graphical checks of Kaplan-Meier curves

One indication that hazards are not proportional can be seen in a Kaplan-Meier plot. If the proportional hazards assumption holds, the curves for each level of the covariate should steadily drift apart. If the estimated survival curves are fairly separated then converge or cross, or if they are initially similar and then sharply diverge, this indicates a violation of proportional hazards.

Example: Bellera et al. (2010) present a review of the proportional hazards assumption with an illustrative example from breast cancer research. They examine 979 women treated for breast cancer with surgery. The endpoint is time until metastasis. Age, tumor size and grade, lymph node involvement, peritumoral vascular invasion (PVI), status of hormone receptors (HRec), Her2 status, and MiB1 expression levels (measure of proliferative activity) were included as covariates.

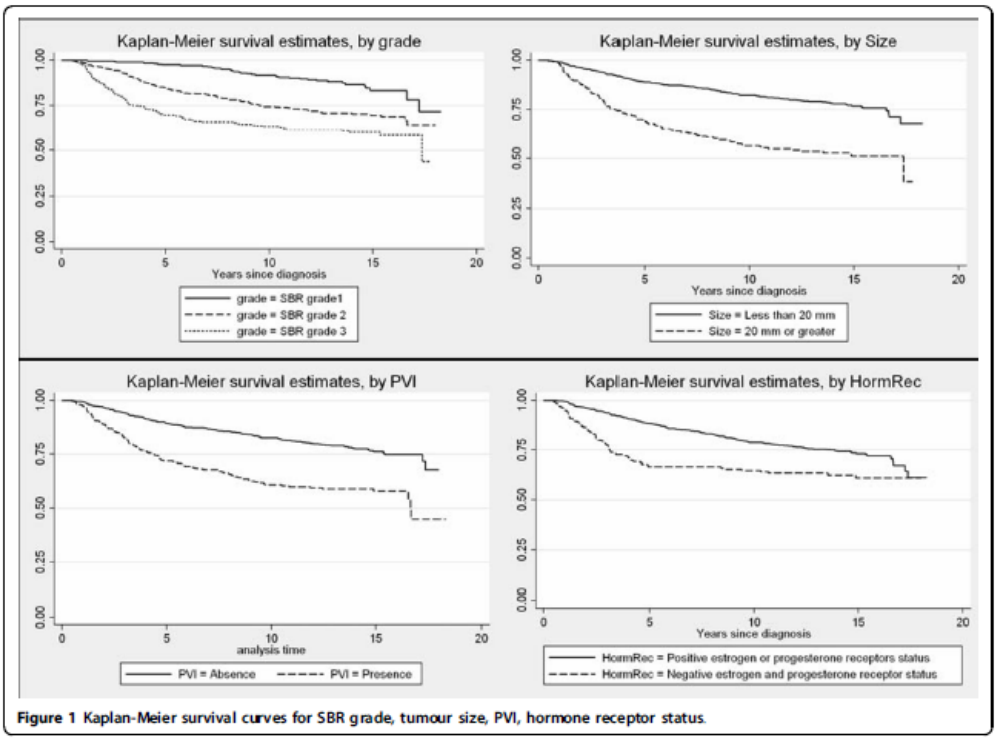


Figure 1 Kaplan-Meier survival curves for SBR grade, tumour size, PVI, hormone receptor status.

Figure 1 above considers the Kaplan-Meier curves for tumor grade (upper left), tumor size divided into two categories (upper right), PVI presence/absence (lower left), and hormone receptor status (positive/negative) (lower right). The Kaplan-Meier curves appear to steadily drift apart for tumor size and PVI, but the curves separate then converge for hormone receptor status and tumor

grade. Thus, the proportional hazards assumption may not be valid for hormone receptor status and tumor grade.

■

Log-log survival curves

We learned previously that, under the proportional hazards model, the log cumulative hazard functions are parallel. To see this again, consider a proportional hazards model with a single binary covariate defining two groups. The survival in one group can be expressed as a function of the survival in the reference group:

$$S_1(t) = [S_0(t)]^{\exp(\beta)}$$

We then apply the **log-log transformation** or **complementary log-log transformation** to both sides. We first take a log transformation, then we take a second log transformation of the negative of the first transformation.

$$\log S_1(t) = \exp(\beta) \log S_0(t)$$
$$\log[-\log S_1(t)] = \beta + \log[-\log S_0(t)]$$

This suggests that, in the presence of proportional hazards, the log of the minus log of the Kaplan-Meier function from different groups will be parallel when plotted against time t (or against log time).

Since the cumulative hazard $H(t) = -\log S(t)$:

$$\log[H_1(t)] = \beta + \log[H_0(t)]$$

Thus, one could achieve the same result by plotting the log of the Nelson-Aalen estimators against time.

Example: Bellera et al. (2010) prepared complementary log-log plots for each of the four covariates considered in Figure 1. Figure 2 considers the log-log survival curves for tumor grade (upper left), tumor size divided into two categories (upper right), PVI (lower left), and hormone receptor status (lower right). For each plot, the y-axis is the log of minus the log of the group-specific survival function $\log(-\log(\hat{S}_i(t))$, which is equivalent to the log cumulative hazard function $\log(\widehat{H}_i(t))$. The x-axis is log time.

Our goal is to visually assess if the group-specific curves are parallel. The curves are roughly parallel for tumor size and PVI. The curves are not parallel for tumor grade and hormone receptor status, indicating a potential violation of the proportional hazards assumption.

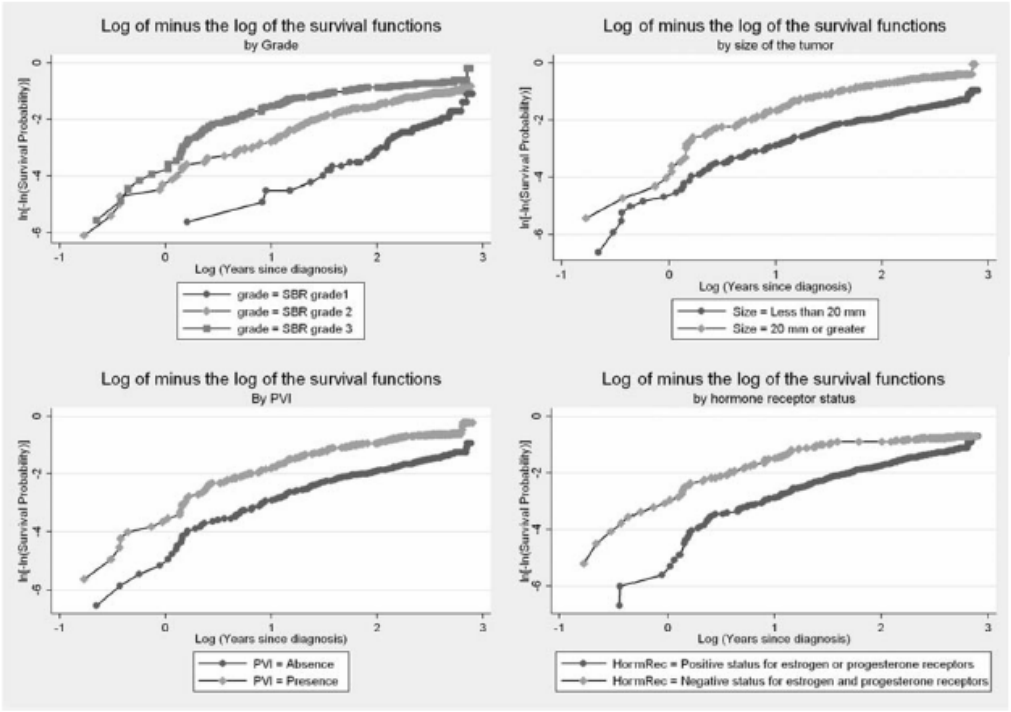


Figure 2 Log(-log(survival)) curves as a function of time (log scale) for SBR grade, tumour size, PVI, hormone receptor status.

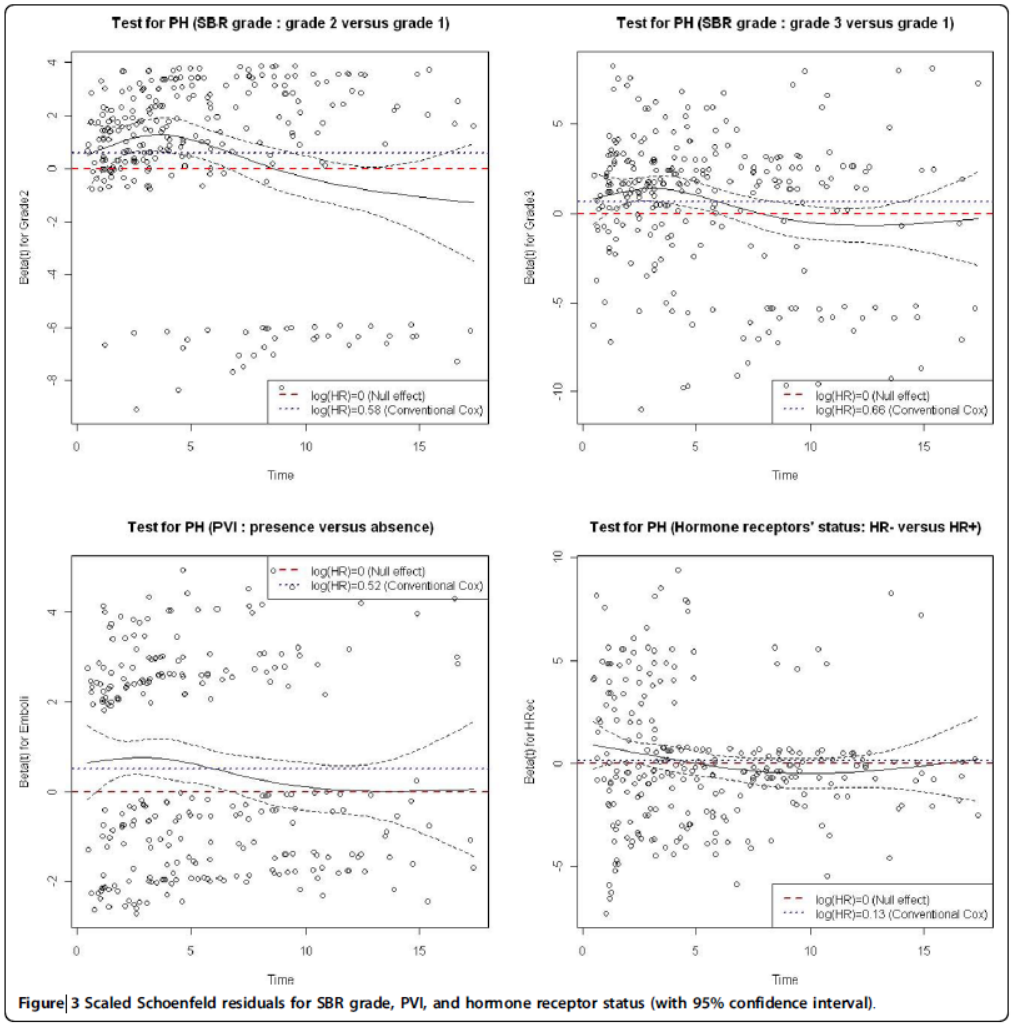
■

These graphical diagnostics are straightforward to implement, but they have limitations. The covariate must be categorical or grouped into categories. The curves are rarely perfectly parallel in practice. The decision to accept the proportional hazards hypothesis often depends on whether these curves cross, which is strong evidence of a proportional hazards assumption violation. The method may be less useful for detecting moderate evidence of proportional hazards assumption violation.

Plotting Schoenfeld residuals

A smooth plot of the scaled Schoenfeld residuals against time can be used to directly visualize whether the log hazard ratio changes over time. If the proportional hazards assumption holds, the log hazard ratio will be constant over time, and it will be well-captured by the fitted coefficient $\hat{\beta}_j$. The Schoenfeld residuals will appear as random scatter above and below a flat line with height $\hat{\beta}_j$, indicating that the log hazard ratio is roughly equal to the fitted coefficient at all times. If proportional hazards is violated, the residuals will be systematically above $\hat{\beta}_j$ at some times, but systematically below $\hat{\beta}_j$ at other times, indicating that the log hazard ratio is not constant. We often fit a smoothed curve (e.g. LOWESS) to the visualize this relationship in the plot.

Example: Bellera et al. (2010) plot the scaled Schoenfeld residuals for several covariates against time. Tumor grade was divided into three categories: 1 (reference), 2 or 3. Thus, there are two coefficients for tumor grade. The scaled Schoenfeld residual plots are shown for tumor grade 2 versus grade 1 (upper left), tumor grade 3 versus grade 1 (upper right), PVI presence versus absence (lower left), and hormone receptor negative versus positive (lower right). A smoothed line with 95% confidence interval is added to each plot. A horizontal dotted line is drawn at the Cox-model estimated log hazard ratio. A red horizontal dashed line is drawn at zero (log hazard ratio under the null H_0).



Figure| 3 Scaled Schoenfeld residuals for SBR grade, PVI, and hormone receptor status (with 95% confidence interval).

The plots for tumor grade suggest a strong effect of grade on survival in the first five years, but this effect seems to diminish over time. Similarly, the impact of PVI tended to decrease over time. Regarding hormone receptor status, the plots suggest that a negative status increased the risk of metastases early on, and became protective afterwards. Thus, the hazard ratio does not seem to be constant over time for any of the variables plotted

■

Testing Schoenfeld residuals slope

Under proportional hazards, the log hazard ratio is constant over time, and the scaled Schoenfeld residuals should have slope 0. We can use a simple linear regression to test if the scaled Schoenfeld residuals have slope different from 0. This test is sometimes referred to as the **Grambsch-Therneau test**. An increasing trend would indicate an increasing hazard ratio over time, and vice versa.

Example: In Bellera et al. (2010), Table 3 reports the p-values from a simple test that the slope of the scaled Schoenfeld residuals is equal to 0. A significant result (e.g. $p<0.05$) indicates evidence of non-proportionality. Tumor grade ($p<0.01$), PVI ($p=0.05$), and hormone receptor status ($p=0.05$) are most likely to contribute to non-proportionality in this model.

Table 3 Test for non-proportionality based on the scaled Schoenfeld residuals from the conventional Cox model (see table 1).

Variable	p-value
Age	0.10
Grade II	<0.01
Grade III	<0.01
Size	0.32
Lymph node involvement	0.22
PVI	0.05
Hormone receptor	0.05
Her2	0.08
Mib1	0.07
GLOBAL	<0.01

■

While this test provides a simple assessment of the proportional hazards assumption, it is advisable to plot the residuals to look for quadratic or logarithmic patterns that may not otherwise be detectable.

Time-by-covariate interactions

With the exception of the test of slope, the above assessments of the proportional hazards assumption are mainly graphical. Another hypothesis-based way to test the proportional hazards assumption is to test for the presence of an interaction between a covariate *and time*. If the proportional hazards assumption holds, the effect of the covariate should be constant with no interaction with time.

Consider a proportional hazards model with two covariates X_{i1} and X_{i2}. The standard proportional hazards model assumes:

$$h_i(t) = h_0(t)\exp(\beta_1 X_{i1} + \beta_2 X_{i2})$$

If we want to test the proportionality of the effect of X_{i2}, we can add an interaction with time:

$$h_i(t) = h_0(t)\exp(\beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i2} Q(t))$$

where $Q(t)$ is some function of time, such as $Q(t) = t$, $Q(t) = \log(t)$, or $Q(t) = \log(t) - \overline{\log(t)}$. In the last version, the interaction is centered at the average of the log survival times $\overline{\log(t)}$. Centering covariates to have mean zero can make it easier to interpret the results.

The hazard ratio for X_{i2} is not constant unless the coefficient for the interaction with time $\beta_3 = 0$. A test of $H_0: \beta_3 = 0$ would be a test of the proportional hazards assumption for X_{i2}.

What if the proportional hazards assumptions fails?

Next week we will talk about ways to relax the proportional hazards assumption, by stratifying the model or allowing time-varying effects.

Part 3. Model Building and Selection

Model Building Review

Suppose that we have a (possibly large) set of covariates in our data. How do we decide which covariates to use in our model?

Consider two scenarios. In the first scenario, we have designed the study to *test a hypothesis of primary interest*. In this case, model building is conducted to adjust the primary hypothesis for other variables (i.e. confounders)

In the second scenario, interest centers upon *identifying a set of variables* that will aid in modeling survival or generating important hypotheses in future study.

In both scenarios, a guiding principle is the simpler, the better. If two models fit the data equally well, then the simpler (smaller) model is preferred.

We also don't want to exceed the limits our data. A commonly adopted rule of thumb is that inferences constructed using proportional hazards regression are considered valid when there are at least 10 events per variable (Peduzzi et al, Journal of Clinical Epidemiology, 1995).

Purposeful selection

In their text book, Hosmer, Lemeshow and May (2008) describe a procedure called purposeful selection of covariates. They remind readers that, while

software shortcuts abound, good model building requires patience and an eye for details. The following is adapted from their description:

Step 1: Fit univariable models for all covariates. Screen for covariates that are significant at the 0.20 to 0.25 level. Fit a multivariable model with all significant covariates and other variables judged to be of clinical importance.

Step 2: Following the fit of the initial multivariable model, use the p-values from the Wald tests of the individual coefficients to identify covariates that might be deleted from the model. For nominal variables, the p-value of a partial likelihood ratio test can be used.

Step 3: Following the fit of the reduced model, assess whether removal of the covariate has produced an "important" change (change by 20% or so) in the coefficients of the remaining variables in the model. If the variable excluded is an important confounder, it should be added back into the model. Continue until no covariates can be deleted.

Step 4: Add to the model, one at a time, all variables excluded from the initial multivariable model to confirm they are neither statistically significant nor an important confounder. This is the preliminary main effects model.

Step 5: Examine the scale of the continuous covariates in the preliminary main effects model. If the variable in the not linear in the log hazard, identify the appropriate transformation. The resulting model is the main effects model.

Step 6: Determine whether interactions are needed. For each plausible interaction term, use the likelihood ratio test to compare the model with and without the interaction. All interactions significant at the 0.05, or other, level are then added jointly to the main effects model.

Step 7: Final model evaluation includes checking for adherence to key model assumptions using diagnostic statistics to check for influential observations and testing for overall goodness-of-fit.

A modification sometimes used in clinical trial settings is to exclude the treatment variable from the variable selection process. Treatment is added to the model *after* creating the preliminary main effects model with other covariates.

Information criteria

Information criteria can also be used to estimate the relative quality of statistical models. Using the same data set, we can imagine fitting many different models, each with a different combination of covariates. We calculate the information criterion for each model. We prefer models with the lowest information criterion.

The **Akaike information criteria (AIC)** can be used to compare models fit to the same data. For Cox regression, the AIC is defined as:

$$AIC = 2p - 2\ell(\widehat{\boldsymbol{\beta}})$$

where p is the number of coefficients in the model and $\ell(\widehat{\boldsymbol{\beta}})$ is value of the log partial likelihood function for that model at the maximum partial likelihood estimate $\widehat{\boldsymbol{\beta}}$.

AIC seeks to balance the trade-off between the goodness-of-fit of the model and the simplicity of the model, also called the model parsimony. Since we seek to reduce the number of coefficients and maximize the likelihood, the model should be chosen to minimize the AIC.

The **Bayesian information criteria (BIC)** is another measure of model quality. Let d be the number of events (failures) in the data set. The BIC for the Cox model is:

$$BIC = p\log(d) - 2\ell(\widehat{\boldsymbol{\beta}})$$

Just as for the AIC, the model should be chosen to minimize the BIC. AIC and BIC can be implemented into a stepwise selection procedure.

Part 4. Looking ahead

Today we continued our discussion of the Cox proportional hazards regression model, focusing on model diagnostics and building. Next week, we will discuss several popular extensions of the Cox proportional hazards model.

So far, we have assumed that all covariates are measured at baseline and do not change over time. In fact, a participant's covariates can change during a study. For example, vaccination halfway through an observational study would be expected to alter someone's hazard of infection with a target disease. This can be handled using **time-dependent covariates**.

We have further assumed that the effect of a covariate is constant over time. In practice, treatments may take time to "kick in," or the effect may wane over time. For example, individuals may develop resistance to HIV treatment. This can be handled using **time-varying effects**.

Under the proportional hazards model, we have one flexible baseline hazard, and the effect of covariates is multiplicative as compared to this single baseline hazard. We can allow for more flexibility if we fit a model with **stratified baseline hazards**. Then, we fit a separate baseline hazard for each stratum.