Use of the Cox Semiparametric Regression Model for Predicting Costs, Charges, and Length of Stay

Frank E. Harrell Jr
L. Richard Smith
Division of Biometry and the Heart Center
Duke University Medical Center
Box 3363 Durham NC 27710 USA
feh@biostat.mc.duke.edu, lrs@biostat.mc.duke.edu

Short Course
Methodologic Issues in Health Services and Outcomes Research
Boston MA

3 December 1995

Contents

Course Philosophy

- Commonly used methods such as linear regression and log–linear regression often do not fit health–care resource consumption data
- A technique that is robust (based on ranks of Y) for modeling regression (case–mix) effects is advantageous
- A technique with fewer distributional assumptions has advantages such as not assuming a mathematical connection between predicted mean and median costs
- Assumptions about transformations of X can be checked using usual regression methods (regression splines, residual plots)
- Graphical techniques coupled with formal statistical tests are the best way to verify model assumptions
- It is frequently best to right–censor costs when costs were truncated because of a bad outcome

Problems with Traditional Ways of Modeling Resource Utilization

1 Multiple Linear Regression

- Y = total hospital costs
- Problems with high outliers $\rightarrow$ too much influence on regression coefficient estimates, etc.
- Often a minimum non–zero cost
- Non–normally distributed residuals $\rightarrow$ improper confidence limits

2 Linear Regression on Transformed Y

- Commonly use $\log(Y)$
- Assumes that patient conditions affect costs multiplicatively
- Residuals still not normal
- Example: hospital charges associated with coronary bypass surgery $\rightarrow$

Had to take logs 6 times to obtain normal distribution

3 Binary Logistic Regression for High Outliers

- Statistically inefficient (lower power, larger s.e.)
- Requires arbitrary choice of high–utilization cutoff
- Does not provide estimate of total system costs

4 Problems with Interrupted Observation

- A hospital with high mortality could have low costs
- Need to penalize when comparing with other hospitals having different mortality probabilities

- Instead of considering a cumulative $12,000 cost at the day of death to be a complete measurement, we could consider the cost to be $12,000+ (right–censored)
- It is hard to unbiasedly estimate the complete cost had the patient lived, but in most cases we will do so more accurately by allowing for censoring rather than ignoring it [1]
- Look at model R^2 with and without censoring

See [6, 20].

[1] See P. 164–166 of [14] for pointers for how to check for informative censoring and to explicitly model the censoring process.

General Regression Models

1 Introduction and Notation

- Regression model using weighted sum of a set of independent or predictor variables
- Interpret parameters and state assumptions by linearizing model with respect to regression coefficients
- Examine regression assumptions

Y	response (dependent) variable
X	$X_1, X_2, \ldots, X_p$ – list of predictors
β	$\beta_0, \beta_1, \ldots, \beta_p$ – regression coefficients
β_0	intercept parameter(optional)
$\beta_1, \ldots, \beta_p$	weights or regression coefficients
$X\beta$	$\beta_0 + \beta_1 X_1 + \ldots + \beta_p X_p, X_0 = 1$

Model: connection between X and Y
$C(Y|X)$: property of distribution of Y given X,

e.g.
$C(Y|X) = \mathrm{E}(Y|X)$ or $\mathrm{Prob}\{Y = 1|X\}$.

2 Model Formulations

General linear regression model

$$C(Y|X) = g(X\beta).$$

Examples

$C(Y|X) =$ $E(Y|X) =$ $X\beta, Y|X \sim n(X\beta, \sigma^2)$
$C(Y|X) = \mathrm{Prob}\{Y = 1|X\} =$ $(1 + \exp(-X\beta))^{-1}$

Linearize: $h(C(Y|X)) = X\beta, h(u) = g^{-1}(u)$
Example:

$$\begin{aligned} C(Y|X) = \mathrm{Prob}\{Y = 1|X\} &= (1 + \exp(-X\beta))^{-1} \\ h(u) = \mathrm{logit}(u) &= \log(\frac{u}{1-u}) \\ h(C(Y|X)) &= C'(Y|X) \text{ (link)} \end{aligned}$$

General linear regression model: $C'(Y|X) = X\beta$.

3 Interpreting Model Parameters

Suppose that X_j is linear and doesn't interact with other X's.

$$\begin{aligned} C'(Y|X) &= X\beta = \beta_0 + \beta_1 X_1 + \ldots + \beta_p X_p \\ \beta_j &= C'(Y|X_1, X_2, \ldots, X_j + 1, \ldots, X_p) \\ &- C'(Y|X_1, X_2, \ldots, X_j, \ldots, X_p) \end{aligned}$$

Drop $'$ from C' and assume $C(Y|X)$ is property of Y that is linearly related to weighted sum of X's.

4 Relaxing Linearity Assumption for Continuous Predictors

4.1 Simple Nonlinear Terms

$$C(Y|X_1) = \beta_0 + \beta_1 X_1 + \beta_2 X_1^2.$$

H_0 : model is linear in X_1 vs. H_a : model is quadratic in $X_1 \equiv H_0 : \beta_2 = 0$.

Polynomials do not adequately fit logarithmic functions or "threshold" effects, and have unwanted peaks and valleys [5].

4.2 Splines for Estimating Shape of Regression Function and Determining Predictor Transformations

Spline Function: piecewise polynomial
Linear Spline Function: piecewise linear function
Ex: X–axis divided into intervals with endpoints a, b, c (knots).

$$f(X) = \beta_0 + \beta_1 X + \beta_2 (X-a)_+ + \beta_3 (X-b)_+ + \beta_4 (X-c)_+,$$

where

$$(u)_+ = \begin{cases} u, & u > 0, \\ 0, & u \leq 0. \end{cases}$$

$$f(X) = \begin{cases} \beta_0 + \beta_1 X, & X \leq a \\ \beta_0 + \beta_1 X + \beta_2 (X-a) & a < X \leq b \\ \beta_0 + \beta_1 X + \beta_2 (X-a) + \beta_3 (X-b) & b < X \leq c \end{cases}$$

$$= \beta_0 + \beta_1 X + \beta_2 (X-a)$$
$$+\beta_3 (X-b) + \beta_4 (X-c) \qquad c < X.$$

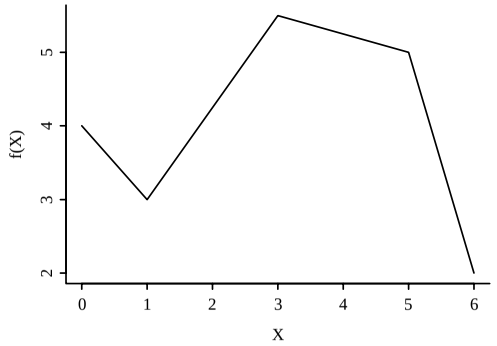

Figure 1: A linear spline function with knots at a=1, b=3, c=5

$$C(Y|X) = f(X) = X\beta,$$

where $X\beta = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4$, and

$$X_1 = X \quad X_2 = (X-a)_+$$
$$X_3 = (X-b)_+ \quad X_4 = (X-c)_+.$$

Overall linearity in X can be tested by testing $H_0 : \beta_2 = \beta_3 = \beta_4 = 0$.

4.3 Cubic Spline Functions

Cubic splines are smooth at knots (function, first, second derivatives agree).

$$\begin{aligned} f(X) &= \beta_0 + \beta_1 X + \beta_2 X^2 + \beta_3 X^3 \\ &+ \beta_4 (X-a)^3_+ + \beta_5 (X-b)^3_+ + \beta_6 (X-c)^3_+ \\ &= X\beta \end{aligned}$$

$$X_1 = X \qquad X_2 = X^2$$
$$X_3 = X^3 \quad X_4 = (X-a)^3_+$$
$$X_5 = (X-b)^3_+ \quad X_6 = (X-c)^3_+.$$

k knots $\rightarrow k+3$ coefficients excluding intercept. See [4, 19, 21] for more information.

4.4 Restricted Cubic Splines

Stone and Koo [24]: cubic splines poorly behaved in tails. Constrain function to be linear in tails. $k+3 \rightarrow k-1$ parameters.

The restricted spline function with k knots $t_1, \ldots, t_k$ is given by

$$f(X) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_{k-1} X_{k-1},$$

where $X_1 = X$ and for $j = 1, \ldots, k-2$,

$$\begin{aligned} X_{j+1} &= (X - t_j)_+^3 - (X - t_{k-1})_+^3 (t_k - t_j)/(t_k - t_{k-1}) \\ &\quad + (X - t_k)_+^3 (t_{k-1} - t_j)/(t_k - t_{k-1}) \end{aligned}$$

(see [5]).

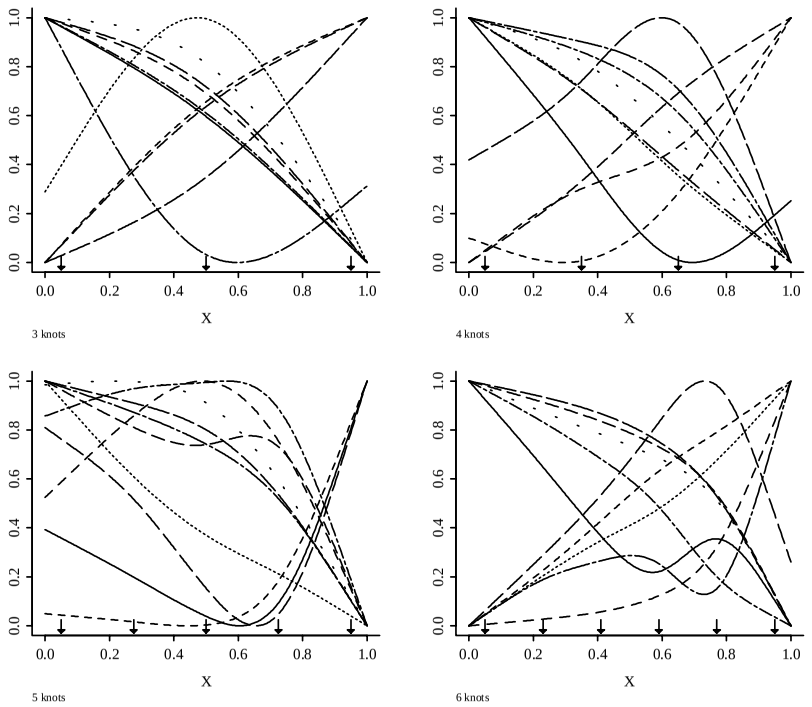

Figure 2: Some typical restricted cubic spline functions for $k = 3, 4, 5, 6$. The y-axis is $X\beta$. Arrows indicate knots.

Once $\beta_0, \ldots, \beta_{k-1}$ are estimated, the restricted cubic spline can be restated in the form

$$\begin{aligned} f(X) &= \beta_0 + \beta_1 X + \beta_2 (X - t_1)_+^3 + \beta_3 (X - t_2)_+^3 \\ &\quad + \ldots + \beta_{k+1} (X - t_k)_+^3 \end{aligned}$$

by computing

$$\begin{aligned}\beta_k &= [\beta_2(t_1 - t_k) + \beta_3(t_2 - t_k) + \ldots + \beta_{k-1}(t_{k-2} - t_k)]/(t_k - t_{k-1})\\ \beta_{k+1} &= [\beta_2(t_1 - t_{k-1}) + \beta_3(t_2 - t_{k-1}) + \ldots + \beta_{k-1}(t_{k-2} - t_{k-1})]/(t_{k-1} - t_k).\end{aligned}$$

A test of linearity in X can be obtained by testing

$$H_0 : \beta_2 = \beta_3 = \ldots = \beta_{k-1} = 0.$$

See [12] for more info.

4.5 Choosing Number and Position of Knots

- Knots are specified in advance in regression splines
- Locations not important in most situations [7, 23]
- Place knots where data exist — fixed quantiles of predictor's marginal distribution
- Fit depends more on choice of k

k	Quantiles						
3			.05	.5	.95		
4			.05	.35	.65	.95	
5		.05	.275	.5	.725	.95	
6	.05	.23	.41	.59	.77	.95	
7	.025	.1833	.3417	.5	.6583	.8167	.975

$n < 100$ – replace outer quantiles with 5th smallest and 5th largest X [24].

Choice of k:

- Flexibility of fit vs. n and variance
- Usually $k = 3, 4, 5$. Often $k = 4$
- Large n (e.g. $n \geq 100$) – $k = 5$
- Small n (< 30, say) – $k = 3$
- Can use Akaike's information criterion (AIC) [2, 26] to choose k
- This chooses k to maximize model likelihood ratio $\chi^2 - 2k$.

5 Steps of One Possible Modeling Strategy

1. Assemble accurate, pertinent data and lots of it.
2. Formulate good hypotheses — specify relevant candidate predictors and possible interactions.

3. Discard observations having missing Y after characterizing
4. Characterize and impute missing X
5. Do data reduction if needed (pre – transformations, combinations), or use penalized estimation [27]
6. Use the entire sample in model development
7. Check linearity assumptions and make transformations in Xs as needed.
8. Check additivity assumptions and add pre–specified interaction terms.
9. Check to see if there are overly–influential observations.
10. Check distributional assumptions and choose a different model if needed.
11. Do limited backwards step–down variable selection if parsimony is more important that accuracy [22].
12. This is the "final" model.
13. Validate this model for calibration and discrimination ability, preferably using bootstrapping.

14. Shrink parameter estimates if there is overfitting but no further data reduction is desired (unless shrinkage built–in to estimation)
15. When missing values were imputed, adjust final variance–covariance matrix for imputation wherever possible
16. When all steps of the modeling strategy can be automated, consider using Faraway's method [9] to penalize for the randomness inherent in the multiple steps.

See [11].

Censored Data

1 Background

- Response variable Y is usually time until an event
- Allow for censoring
- Ex: 5y follow–up study; subject still alive at 5y has failure time 5+
- Length of follow–up can vary
- Response variable can actually be anything
- Must usually have independent censoring: Random variable representing response is statistically independent of random variable representing censoring value.
  Subjects are not selectively censored when they appear to be at a low or high risk of the event of interest.
- Minimal assumption: non–informative censoring.

Parameters of censoring distribution do not overlap with parameters of response distribution.
Likelihood function separates into two components that can be maximized separately.

2 Notation, Survival and Hazard Functions

$$S(y) = \text{Prob}\{Y > y\} = 1 - F(y)$$

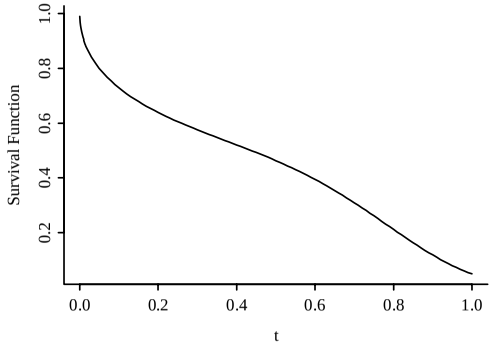

Figure 3: Survival function

- Hazard function (force of mortality; instantaneous event rate)

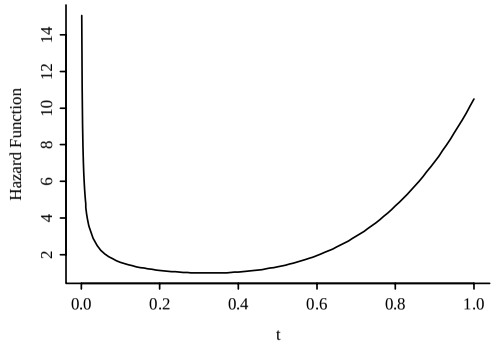

Figure 4: Hazard function

- Y discrete $\rightarrow$

$$\lambda(y) = \text{Prob}\{Y = y | Y \geq y\},$$

which using the law of conditional probability becomes

$$\begin{aligned} \lambda(t) &= \text{Prob}\{Y = y\}/\text{Prob}\{Y \geq y\} \\ &= \frac{f(y)}{S(y)}, \end{aligned}$$

- $f(y)$ is the probability density function of Y evaluated at y: the derivative or slope of the cumulative distribution function $1 - S(y)$.
- Quantiles and mean of distribution of Y:

$$Y_q = S^{-1}(1-q)$$

$$\begin{aligned} Y_{0.5} &= S^{-1}(0.5) \\ \mu &= \int_0^\infty S(v)dv \quad (Y+) \end{aligned}$$

- Potential response for subject i: Y_i
- Censoring value of response: D_i
- Event indicator:

$$\begin{aligned} e_i &= 1 \text{ if the event was observed } (Y_i \leq D_i), \\ &= 0 \text{ if the response was censored } (Y_i > D_i). \end{aligned}$$

- The observed response is

$$y_i = \min(Y_i, D_i),$$

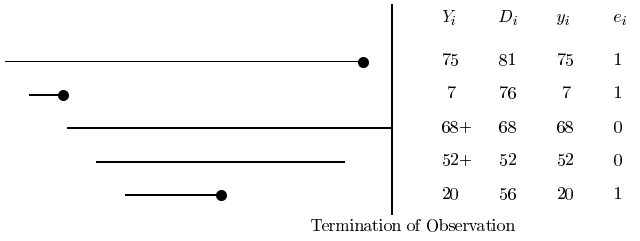

Figure 5: Some censored data. Circles denote complete response observed.

3 Homogeneous Distributions (No Case–Mix Adjustment)

4 Nonparametric Estimation of S

4.1 Kaplan–Meier Estimator

- No censoring $\rightarrow$

$$S_n(y) = [\text{number of } Y_i > y]/n.$$

- Kaplan–Meier (product–limit) estimator

y	No. Subjects At Risk	Complete	Censored	Cumulative Probability
12	100	1	0	$99/100 = .99$
30	99	2	1	$97/99 \times 99/100 = .97$
60	96	0	3	$96/96 \times .97 = .97$
72	93	3	0	$90/93 \times .97 = .94$
.	.	.	.	.
.	.	.	.	.

$$S_{\text{KM}}(y) = \prod_{i:y_i<y} (1 - c_i/n_i),$$

c_i = number of complete responses at y_i.

- Simple example

$$1 \quad 3 \quad 3 \quad 6^+ \quad 8^+ \quad 9 \quad 10^+.$$

i	y_i	n_i	c_i	$(n_i - c_i)/n_i$
1	1	7	1	6/7
2	3	6	2	4/6
3	9	2	1	1/2

$$\begin{aligned} S_{\text{KM}}(y) &= 1, \quad 0 \le y < 1 \\ &= 6/7 = .85, \quad 1 \le y < 3 \\ &= (6/7)(4/6) = .57, \quad 3 \le y < 9 \\ &= (6/7)(4/6)(1/2) = .29, \quad 9 \le y < 10. \end{aligned}$$

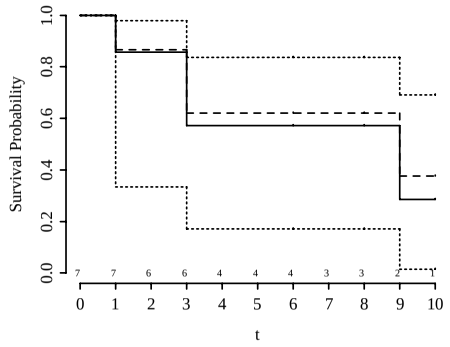

Figure 6: Kaplan–Meier product-limit estimator with 0.95 confidence bands. The Altschuler–Nelson–Fleming–Harrington estimator is depicted with the dashed lines.

Proportional Hazards Regression Model

1 Allowing for Covariables through Multiplicative Hazards Effects

$$\begin{aligned}
\lambda(y|X) &= \lambda(y)\exp(X\beta) \\
S(y|X) &= \exp[-\Lambda(y)\exp(X\beta)] = \exp[-\Lambda(y)]^{\exp(X\beta)} \\
\Lambda(y) &= \int_0^y \lambda(u)du \\
S(y|X) &= S(y)^{\exp(X\beta)}
\end{aligned}$$

1.1 Model Assumptions and Interpretation of Parameters

$$\begin{aligned}
\log\lambda(y|X) &= \log\lambda(y) + X\beta \\
\log-\log S(y|X) &= \log-\log S(y) + X\beta.
\end{aligned}$$

Assumptions:

- Linear effect of predictors on $\log\lambda$, $\log\Lambda$ $\log-\log S$

- No interaction between X and $y \rightarrow$ impact of X same over response values

$$\begin{aligned}\beta_j &= \log\lambda(y|X_1, X_2, \ldots, X_j+1, X_{j+1}, \ldots, X_k) \\ &- \log\lambda(y|X_1, \ldots, X_j, \ldots, X_k) \\ &= \log - \log S(y|X_1, X_2, \ldots, X_j+1, X_{j+1}, \ldots, X_k) \\ &- \log - \log S(y|X_1, \ldots, X_j, \ldots, X_k)\end{aligned}$$

- Effect of increasing X_j by d is to increase λ by factor of $\exp(\beta_j d)$ or to raise $S(y)$ to the power $\exp(\beta_j d)$ or to increase $\log - \log S(y)$ by $\beta_j d$.

1.2 Assessment of Model Fit

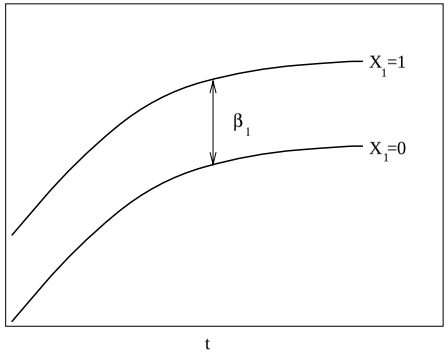

Figure 7: PH Model with one binary predictor. Y–axis is $\log\lambda(y)$ *or* $\log\Lambda(y)$*. For* $\log\Lambda(y)$*, the curves must be non–decreasing. For* $\log\lambda(y)$*, they may be any shape.*

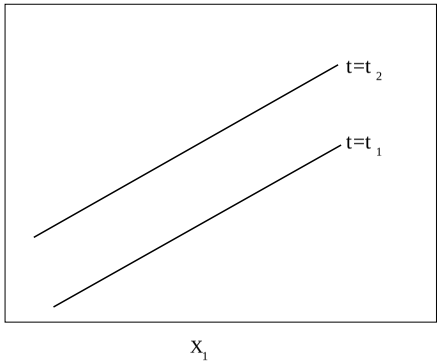

Figure 8: PH model with one continuous predictor. Y–axis is $\log\lambda(y)$ *or* $\log\Lambda(y)$*. For* $\log\Lambda(y)$*, drawn for* $y_2 > y_1$*. The slope of each line is* β_1*.*

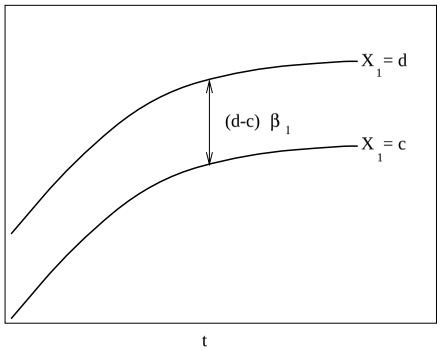

Figure 9: PH model with one continuous predictor. Y–axis is $\log\lambda(y)$ *or* $\log\Lambda(y)$*. For* $\log\lambda$*, the functions need not be monotonic.*

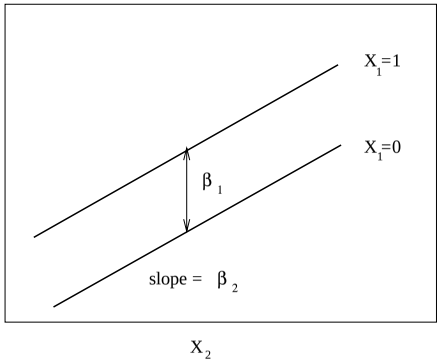

Figure 10: Regression assumptions, linear additive PH model with two predictors. Y–axis is $\log \lambda(y)$ or $\log \Lambda(y)$ for a fixed y.

2 Cox Model

2.1 Preliminaries

- Developed by DR Cox [3]
- Most popular survival model
- Semi–parametric (non–parametric hazard; parametric regression)
- Usually more interest in effects of X than on shape of $\lambda(y)$
- Uses only rank ordering of responses $\rightarrow$ more robust

- Even if parametric PH assumptions true, Cox model still fully efficient for β
- Model diagnostics are advanced

2.2 Model Definition

$$\begin{aligned} \lambda(y|X) &= \lambda(y)\exp(X\beta) \\ S(y|X) &= S(y)^{\exp(X\beta)} \end{aligned}$$

- No intercept parameter
- No assumption about shape of λ or S
- Does not assume a simple connection between mean and median
- All monotonic transformations of Y yield same $\hat{\beta}$

2.3 Extending the Model by Stratification

- Is a unique feature of the Cox model
- Adjust for non–modeled factors
- Factors too difficult to model or fail PH assumption

- Commonly used to adjust for variation across hospitals
- Allow form of λ to vary across strata
- Rank responses within strata
- Stratum ID is C

$$\begin{aligned}\lambda(y|X, C = j) &= \lambda_j(y)\exp(X\beta), \quad \text{or}\\ S(y|X, C = j) &= S_j(y)^{\exp(X\beta)}.\end{aligned}$$

- Not assume connection between shapes of λ_j
- By default, assume common β
- Ex: model age, stratify on sex
  Estimates common age slope pooling F and M
  No assumption about effect of sex except no age interact.
- Can stratify on multiple factors (cross–classify)
- Loss of efficiency not bad unless number of events in strata very small
- Stratum with no events is ignored

- Estimate β by getting separate log–likelihood for each stratum and adding up (independence)
- No inference about strat. factors
- Useful for checking PH and linearity assumptions: Model, then stratify on an X
- Can extend to strata $\times$ covariable interaction

$$\begin{aligned}\lambda(y|X_1, C = 1) &= \lambda_1(y)\exp(\beta_1 X_1)\\ \lambda(y|X_1, C = 2) &= \lambda_2(y)\exp(\beta_1 X_1 + \beta_2 X_1).\end{aligned}$$

$$\lambda(y|X_1, C = j) = \lambda_j(y)\exp(\beta_1 X_1 + \beta_2 X_2)$$

- X_2 is product interaction term (0 for F, X_1 for M)
- Are testing interaction with sex without modeling main effect!

3 Estimation of β

- Cox partial likelihood

- If no ties in Y Is a marginal likelihood of the ranks of responses
- Several methods for handling tied Y; Efron's [8] is a good default
- For heavy ties (e.g., some length of stay studies), may need to handle ties exactly (SAS PROC PHREG does this efficiently) or break ties by adding small random errors to Y

4 Estimation of Survival Probability and Secondary Parameters

- Kalbfleisch–Prentice discrete hazard model method $\rightarrow$ K–M if $\hat{\beta} = 0$

$$\hat{S}(y|X) = \hat{S}(y)^{\exp(X\hat{\beta})}.$$

- Stratified model $\rightarrow$ estimate underlying hazard parameters separately within strata
- "Adjusted K–M estimates"
- Use to estimate quantiles and (if largest response uncensored) the mean

- For mean, compute area under step–function:

$$\begin{aligned}\mu_X &= y_1\hat{S}(y_1)^{\exp(X\hat{\beta})} + (y_2 - y_1)\hat{S}(y_2)^{\exp(X\hat{\beta})} \\ &\quad + \ldots + (y_k - y_{k-1})\hat{S}(y_k),\end{aligned}$$

  where the unique uncensored responses are $y_1, \ldots, y_k$
- Highest response censored $\rightarrow$ can compute mean restricted cost
- No censoring, no covariables $\rightarrow$ reproduces $\bar{Y}$
- Computational trick for estimating mean Y for many different subjects:
  Compute areas under $\hat{S}(y|X)$ once for a sequence of $X\hat{\beta}$
  Save the areas, and for any new $X\hat{\beta}$ use linear interpolation on this sequence to estimate new mean
- Determine relationship between hazard ratios and mean or median cost ratios by plotting $X_i\hat{\beta}$ vs. predicted mean or median cost given X_i

5 Residuals

Residual	Purposes
martingale	assessing adequacy of a hypothesized predictor transformation
	graphing an estimate of a predictor transformation (Section 6.1)
Schoenfeld	testing PH assumption (Section 6.2)
	graphing estimate of hazard ratio function (Section 6.2)

6 Assessment of Model Fit

6.1 Regression Assumptions

Example: A 4–knot spline Cox PH model in two variables (X_1, X_2) which assumes linearity in X_1 and no $X_1 \times X_2$ interaction

$$\begin{aligned}\lambda(y|X) &= \lambda(y)\exp(\beta_1 X_1 + \beta_2 X_2 + \beta_3 X_2' + \beta_4 X_2''),\\ &= \lambda(y)\exp(\beta_1 X_1 + f(X_2)),\end{aligned}$$

$$f(X_2) = \beta_2 X_2 + \beta_3 X_2' + \beta_4 X_2''.$$

$$\log\lambda(y|X) = \log\lambda(y) + \beta_1 X_1 + f(X_2).$$

To not assume PH in X_1, stratify on it:

$$\begin{aligned}\log\lambda(y|X_2, C=j) &= \log\lambda_j(y) + \beta_1 X_2 + \beta_2 X_2'\\ &\quad + \beta_3 X_2''\\ &= \log\lambda_j(y) + f(X_2).\end{aligned}$$

- Example of modeling a single continuous variable (left ventricular ejection fraction), response = time to cardiovascular death
  The AICs for 3, 4, 5, and 6–knots spline fits were respectively 126, 124, 122, and 120.

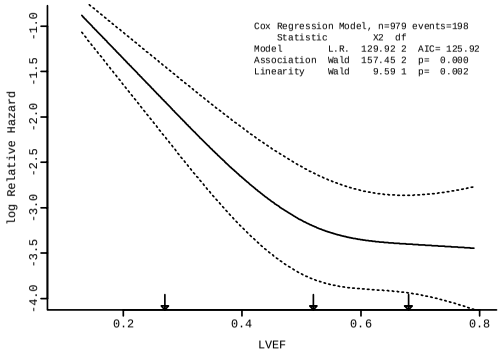

Figure 11: Restricted cubic spline estimate of relationship between LVEF and relative log hazard from a sample of 979 patients and 198 cardiovascular deaths. Data from the Duke Cardiovascular Disease Databank.

Smoothed residual plot: Martingale residuals, loess smoother

- One vector of residuals no matter how many covariables
- Unadjusted estimates of regression shape obtained by fixing $\hat{\beta} = 0$ for all Xs

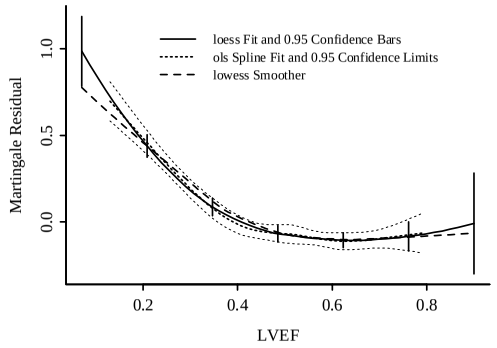

Figure 12: Three smoothed estimates relating martingale residuals [25] to LVEF.

Purpose	Method
Estimate transformation for a single variable	Force $\hat{\beta}_1 = 0$ and compute residuals off of the null regression
Check linearity assumption for a single variable	Compute $\hat{\beta}_1$ and compute residuals off of the linear regression
Estimate marginal transformations for p variables	Force $\hat{\beta}_1, \ldots, \hat{\beta}_p = 0$ and compute residuals off the global null model
Estimate transformation for variable i adjusted for other $p-1$	Estimate $p-1$ βs, forcing $\hat{\beta}_i = 0$, compute residuals off of mixed global/null model

6.2 Proportional Hazards Assumption

- Parallelism of $\log - \log S(y)$ plots
- Comparison of stratified and modeled estimates of $S(y)$
- Stratify Y, get interval–specific Cox regression coefficients:
  In an interval, exclude all subjects with response before start of interval
  Censor all events at end of interval

Example:

Response Interval	Observations	Complete Response	Log Hazard Ratio	Standard Error
$[0, 209)$	40	12	-0.47	0.59
$[209, 234)$	27	12	-0.72	0.58
234+	14	12	-0.50	0.64

Overall Cox $\hat{\beta} = -0.57$.

- Schoenfeld residuals r computed at each unique uncensored y
- Partial derivative of $\log L$ with respect to each X in turn

- Grambsch and Therneau scale to yield estimates of $\beta(y)$: $\hat{\beta} + dr\hat{V}$, d=no. uncensored responses
- Can form a powerful test of PH (Z:PH in old SAS PROC PHGLM)

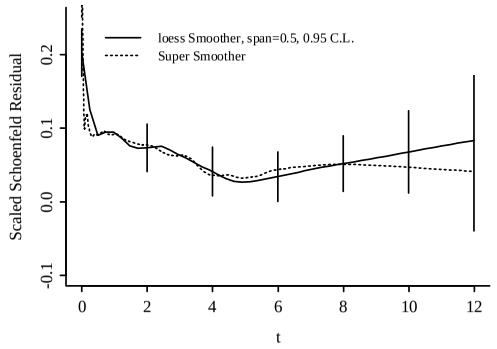

Figure 13: Smoothed weighted [10] Schoenfeld [18] residuals. Test for PH based on the correlation (ρ) between the individual weighted Schoenfeld residuals and the rank of response yielded $\rho = -0.23, z = -6.73, P = 2 \times 10^{-11}$.

- Can test PH by testing $y \times X$ interaction using time– dependent covariables

Assumptions of the Proportional Hazards Model

$$\lambda(y|X) = \lambda(y)e^{\beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_p X_p}$$

Variables	Assumptions	Verification
Response Variable Y	Shape of $\lambda(y\|X)$ for fixed X as $y \uparrow$ Cox: none Weibull: y^θ	Shape of $S_{\text{KM}}(y)$
Interaction between X and Y	Proportional hazards — effect of X does not depend on Y. E.g. treatment effect is constant over y cutoff.	• Categorical X: check parallelism of stratified $\log[-\log S(y)]$ plots as $y \uparrow$ • Muenz [15] cum. hazard ratio plots • Arjas [1] cum. hazard plots • Check agreement of stratified and modeled estimates • Hazard ratio plots • Smoothed Schoenfeld residual plots and correlation test (y vs. residual) • Test time–dependent covariable such as $X \times \log(y+1)$ • Ratio of parametrically estimated $\lambda(y)$
Individual Predictors X	Shape of $\lambda(y\|X)$ for fixed y as $X \uparrow$ Linear: $\log\lambda(y\|X) = \log\lambda(y) + \beta X$ Nonlinear: $\log\lambda(y\|X) = \log\lambda(y) + f(X)$	• k–level ordinal X : linear term + $k-2$ dummy variables • Continuous X: Polynomials, spline functions, smoothed martingale residual plots
Interaction between X_1 and X_2	Additive effects: effect of X_1 on $\log\lambda$ is independent of X_2 and vice–versa	Test non–additive terms, e.g. products

Method	Requires Grouping X	Requires Grouping y	Computational Efficiency	Yields Formal Test	Yields Estimate of $\lambda_2(y)/\lambda_1(y)$	Requires Fitting 2 Models	Must Choose Smoothing Parameter
log[−log], Muenz, Arjas plots	x		x			x	
Dabrowska log $\hat{\Lambda}$ difference plots	x		x	x		x	
Stratified vs. Modeled Estimates	x		x			x	
Hazard ratio plot		x		?	x	x	?
Schoenfeld residual plot			x		x		x
Schoenfeld residual correlation test			x	x			
Fit time–dependent covariables				x	x		
Ratio of parametric estimates of $\lambda(y)$	x		x	x	x	x	

See Hess [13] for an excellent review of graphical methods for assessing PH.

7 What to Do When PH Fails

- Test of association not needed → stratify
- P–value for testing variable may still be useful (conservative)
- Distribution estimates wrong in certain intervals of y
- Can model non–PH:

$$\lambda(y|X) = \lambda_0(y)\exp(\beta_1 X + \beta_2 X \times \log(y+1))$$

- Can also use response intervals:

$$\lambda(y|X) = \lambda_0(y)\exp[\beta_1 X + \beta_2 X \times I(y > c)],$$

8 Quantifying Predictive Ability

- $R^2_{\text{LR}} = 1 - \exp(-\text{LR}/n)$ [16]
  Divide by max attainable value to get R^2_{N}.
- c = concordance probability (predicted vs. observed)
- Is a generalized ROC area
- All possible pairs of subjects whose ordering of responses can be determined
- Fraction of these for which X ordered same as Y
- Somers' $D_{xy} = 2(c - 0.5)$

9 Validation of Discrimination and Other Statistical Indexes

Validate R^2, D_{xy} indexes, optimally using the bootstrap so that don't hold back data [11].

Can also validate slope calibration to estimate shrinkage from overfitting:

$$\lambda(y|X) = \lambda(y)\exp(\gamma X b).$$

Example in which all predictors are noise:

Index	Final Model Fit	Bootstrap Sample Models	Evaluate Orig. Sample	Optimism	Corrected Index
D_{xy}	-0.16	-0.31	-0.09	-0.22	0.06
R^2_{N}	0.05	0.15	0.00	0.15	-0.10
Slope	1.00	1.00	0.25	0.75	0.25

10 Describing the Fitted Model

- Can use coefficients if linear and additive
- In general, use e.g. inter–quartile–range hazard ratios for various levels of interacting factors
- Translate to cost ratios
- Nomogram to compute $X\hat{\beta}$
- Also $\hat{S}(y|X)$ for a few y
- Axis for median cost

- Axis for mean cost

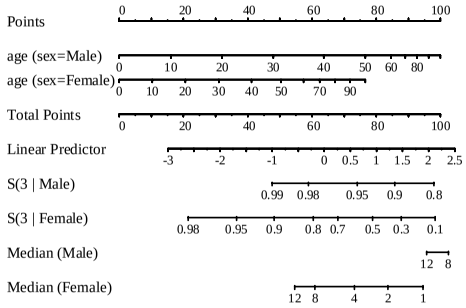

Figure 14: Nomogram from a fitted stratified Cox model that allowed for interaction between age and sex, and nonlinearity in age. The axis for median survival time is truncated on the left where the median is beyond the last follow-up time.

Case Study

- Ambulatory care — family medicine clinic
- Data courtesy of George R Parkerson [17]
- 413 patients from Caswell Family Medical Center, NC
- Y = total charges for office health care during 18–month follow–up
- No censored charges
- 106 patients have $Y = 0$
- Median=$99, mean=$181, 75%=$260, 95%=$609, 99%=$1202
- Median non–zero charge=$157
- Predictors:

Variable	Meaning
hyperten	hypertensive vs. normotensive (HT, NT) $n = 116$ vs. 297
age	
sex	
dusoi	Duke U. severity of illness checklist (0–100)
perceive	Perceived health status (0, 50, 100)
dis	Perceived disability (0, 50, 100)
numdx	Number of diagnoses (1–6)

- `perceive`, `dis` are from the Duke Health Profile
- Distribution of 18m charges

```
# S-PLUS commands

hist(charge, nclass=30, xlab='Total Charges, $')
```

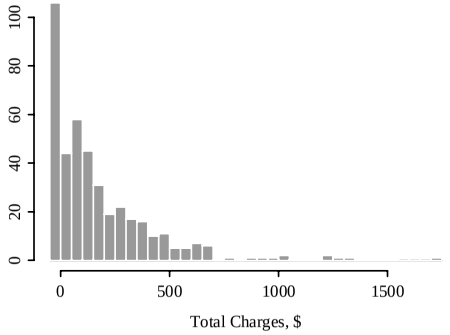

Figure 15: Histogram of total charges

- One minus cumulative dist. of charges, by `hyperten`

```
# Using UNIX S-PLUS Design library in statlib in conjunction
# with Terry Therneau's survival4 package (in statlib or S-PLUS 3.3)

S ← Surv(charge)   # convert to survival time variable
f ← survfit(S ~ hyperten, conf.type='none')
survplot(f, label.curves='equal', ylab='1 - Cumulative Distribution Function')
abline(h=.5, lty=2)
abline(v=tapply(charge,hyperten,median), lty=2)
text(1000,.75,paste('Means=',paste(
     format(round(tapply(charge,hyperten,mean),1)),collapse=' '),sep=''))
title('Survival Distribution of Charges by HT')
```

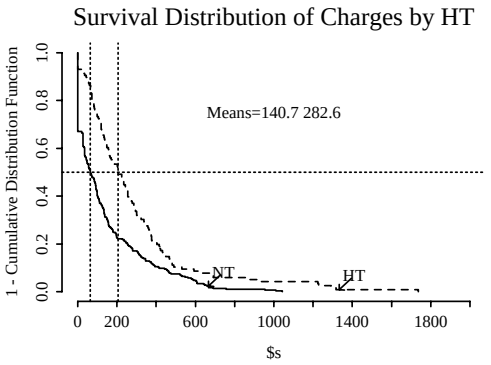

Figure 16: $\hat{S}(y)$ stratified by hypertension

- log – log plots for four important X's

```
par(mfrow=c(2,2))
f ← survfit(S ~ sex, conf.type="none")
survplot(f, loglog=T, logt=T, xlab='log $')
title('Sex')

f ← survfit(S ~ hyperten, conf.type="none")
survplot(f, loglog=T, logt=T, xlab='log $')
title('Hypertension')

f ← survfit(S ~ cut2(dis,50), conf.type="none")
survplot(f, loglog=T, logt=T, xlab='log $')
title('Disability Scale')

f ← survfit(S ~ cut2(dusoi,g=2), conf.type="none")
# g=2 : cut at median dusoi
survplot(f, loglog=T, logt=T, xlab='log $')
title('DUSOI')
```


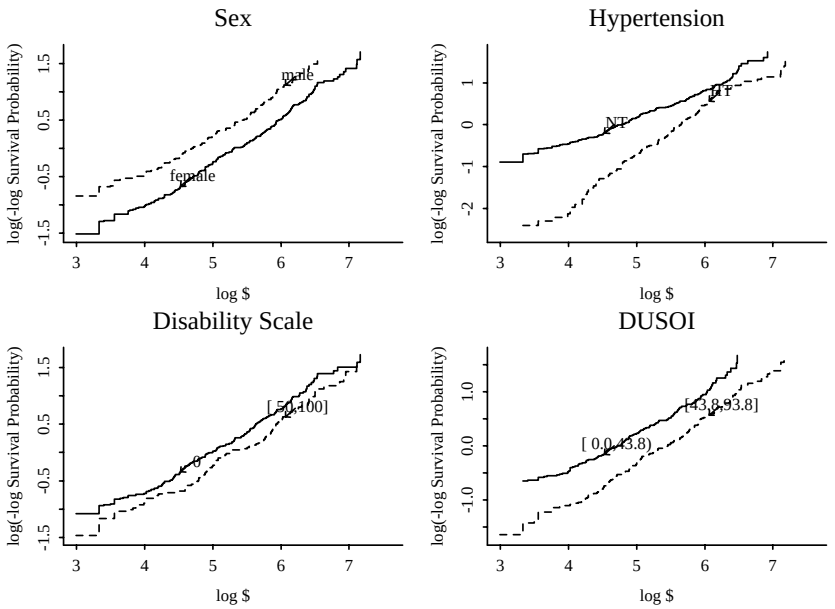

Figure 17: Log-log Kaplan–Meier estimates vs. log y

- Prepare for Cox modeling: Handle ties by adding a random charge between 0 and 1$ for each 0 charge
- Fit preliminary Cox model allowing continuous variables to behave nonlinearly in log hazard
- Use restricted cubic splines with 4 knots

```
set.seed(19)   # set random number seed so can reproduce results
charge ← ifelse(charge==0, runif(length(charge),0,1), charge)
```

```
S ← Surv(charge)

f ← cph(S ~ rcs(age,4) + sex + rcs(dusoi,4) + hyperten + perceive + dis +
        pol(numdx,2))
anova(f)       #actually used latex(anova(f))
```

Table 1: *Wald Statistics for* S

	χ^2	*d.f.*	*P*
age	5.64	3	0.1303
Nonlinear	0.45	2	0.7976
sex	22.57	1	< 0.0001
dusoi	8.38	3	0.0388
Nonlinear	0.28	2	0.8696
hyperten	14.31	1	0.0002
perceive	4.46	1	0.0347
dis	13.39	1	0.0003
numdx	5.80	2	0.0551
Nonlinear	0.34	1	0.5576
TOTAL NONLINEAR	1.15	5	0.9493
TOTAL	112.75	12	< 0.0001

- Go to linear model and check proportional hazards assumption

```
f ← cph(S ~ age + sex + dusoi + hyperten + perceive + dis + numdx,
        surv=T, type="kaplan-meier", x=T, y=T)

z ← cox.zph(f, trans=log)
z

par(mfrow=c(3,3))
plot(z)
```



```
                rho  chisq      p
        age   0.022   0.18 0.6672
    sex=male -0.012   0.07 0.7979
      dusoi   0.092   3.45 0.0632
hyperten=HT   0.124   6.22 0.0127
   perceive -0.009   0.03 0.8574
        dis   0.038   0.62 0.4306
      numdx   0.000   0.00 0.9926
     GLOBAL      NA  14.60 0.0416
```

- Overall test of PH: $P = 0.04$
- Culprit is hyperten ($P = 0.01$)

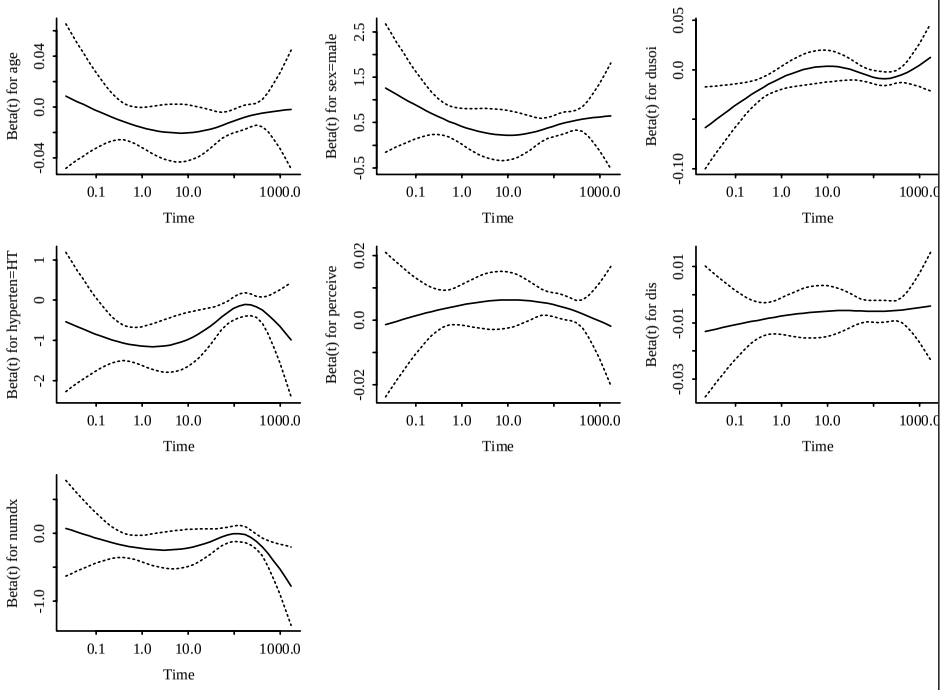

Figure 18: Scaled Schoenfeld residuals. Trend for the `HT` *effect is significantly non-flat ($P = 0.0127$), indicating non-PH.* `dusoi` *had* $P = 0.06$.

- Stratify on hyperten to allow non–PH

```
f ← cph(S ∼ strat(hyperten) + age + sex + dusoi + perceive + dis +
        numdx, surv=T, type='kaplan-meier', x=T, y=T, time.inc=100)
f
z ← cox.zph(f, trans=log)
z
```

```
Obs Events Model L.R. d.f. P Score Score P    R2
413    413      86.16    6 0 81.81       0 0.188

           No Event Event
hyperten=NT        0   297
```



```
hyperten=HT        0   116

             coef se(coef)     z        p
     age -0.00885  0.00425 -2.08 3.75e-02
sex=male  0.46626  0.10551  4.42 9.91e-06
   dusoi -0.00900  0.00309 -2.92 3.55e-03
perceive  0.00398  0.00169  2.36 1.85e-02
     dis -0.00658  0.00177 -3.71 2.06e-04
   numdx -0.10245  0.05273 -1.94 5.20e-02

            rho  chisq      p
     age  0.021   0.17  0.678
sex=male -0.011   0.05  0.816
   dusoi  0.079   2.54  0.111
perceive  0.007   0.02  0.883
     dis  0.014   0.08  0.776
   numdx  0.013   0.06  0.800
  GLOBAL     NA   4.03  0.673
```

- Display model in mathematical form

```
latex(f)    # latex uses print.display package from statlib
```

$$\text{Prob}\{T \geq t \mid \text{hyperten} = i\} = S_i(t)^{e^{X\beta}}, \quad \text{where}$$

$X\hat{\beta} =$

$0.6016 - 0.00885\,\text{age}$

$+0.4663\{\text{male}\} - 0.008999\,\text{dusoi} + 0.003977\,\text{perceive} - 0.006579\,\text{dis} - 0.1025\,\text{numdx}$

and $\{c\} = 1$ if subject is in group c, 0 otherwise.

t	$S_{NT}(t)$	$S_{HT}(t)$
0	1.000	1.000
100	0.432	0.699
200	0.228	0.486
300	0.162	0.258
400	0.082	0.145
500	0.048	0.055
600	0.022	0.042
700	0.001	0.024
800	0.001	0.024
900	0.001	0.019
1000	0.000	0.014
1100	0.000	0.014
1200	0.000	0.014
1300	0.000	0.003
1400	0.000	0.001
1500	0.000	0.001
1600	0.000	0.001
1700	0.000	0.001

- Compute hazard ratio estimates
- For continuous var. use IQR ratios

```
plot(summary(f))
```


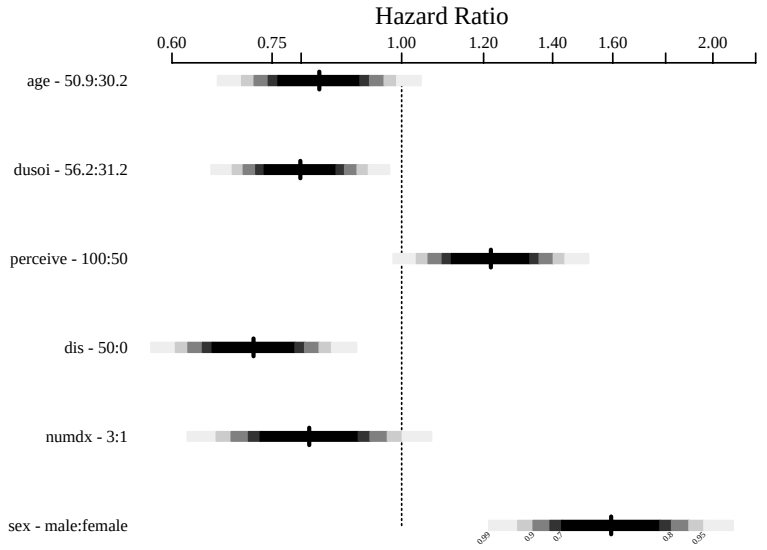

Figure 19: Estimated hazard ratios. Outer quartiles used for non–binary variables

- Now use UNIX S–PLUS `Design` library functions to get the estimated mean and median charges
- The `Mean` function when operating on a Cox model fit returns another S–PLUS function to compute the estimated mean on demand, as a function of $X\hat{\beta}$ and the stratum number

```
mean.charge ← Mean(f, method='approximate')
mean.charge    # prints function definition
```

```
#lp.seq is sequence of linear predictor values for which areas under
#curve were computed
function(lp = 0, stratum = 1, lp.seq = c(-1.742, -1.70241, -1.66283, -1.62324,
    . . . .
1.14763, 1.18722), areas = list("hyperten=NT" = c(563.455, 551.992,
    . . . .
22.7064), "hyperten=HT" = c(944.281, 924.2, 903.982, 883.648, 863.221,
    . . . .
98.2788, 94.9573, 91.7319, 88.6001, 85.5592, 82.6071)))
{
if(length(stratum) > 1)
stop("does not handle vector stratum")
area ← areas[[stratum]]
if(length(lp.seq) == 1 && all(lp == lp.seq))
ymean ← rep(area, length(lp))
else ymean ← approx(lp.seq, area, xout = lp)$y  #linear interpolation
if(any(is.na(ymean)))
warning("means requested for linear predictor values outside range of
linear predictor values in original fit"
)
names(ymean) ← names(lp)
ymean
}
```

```
#Pull off sequence of X*Beta hats used
lp ← mean.charge$lp.seq

#Pull off areas under survival curve estimates for first stratum (no hyperten)
mean.no ← mean.charge$areas[[1]]

#Now do it for second stratum
mean.yes← mean.charge$areas[[2]]
```

- Plot $X\hat{\beta}$ vs. mean, separately by strata. Then do for median.

```
plot(lp, mean.no, type='l',
     xlab='X*Beta', ylab='Predicted Charge', ylim=c(0,950))
lines(lp, mean.yes, lty=2)

quan ← Quantile(f)  # composes function to compute medians on demand
median.no ← quan(lp=lp, stratum=1)
median.yes← quan(lp=lp, stratum=2)

lines(lp, median.no)
lines(lp, median.yes, lty=2)
text(c(-.62,-.596,-.548,-.43),c(129,215,324,416),
     c('median','mean','median','mean'), srt=-25)
```


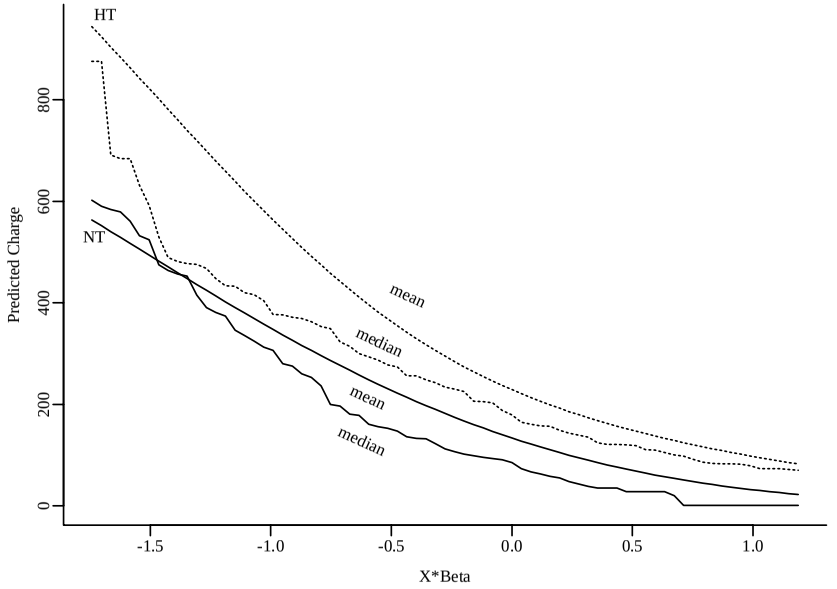

Figure 20: $X\hat{\beta}$ vs. predicted mean and median charges

- Plot predicted mean vs. predicted median charges

```
plot(median.no, mean.no, type='l', xlab='Predicted Median',
     ylab='Predicted Mean', xlim=range(c(median.no,median.yes)),
     ylim=range(c(mean.no,mean.yes)))
lines(median.yes, mean.yes, lty=2)
```

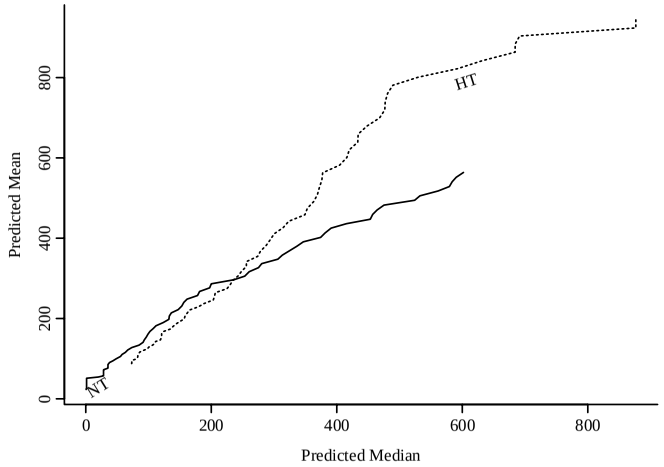

Figure 21: Relationship between predicted mean and median charges, by strata

- Convert hazard ratios to mean or median charge ratios

```
# Solve for mean charge in no hypertension stratum where X*beta=0
mean0.no    ← approx(lp, mean.no, xout=0)$y

mean0.yes   ← approx(lp, mean.yes,xout=0)$y
median0.no  ← approx(lp, median.no, xout=0)$y
median0.yes ← approx(lp, median.yes, xout=0)$y

elp ← exp(lp)
plot(elp, mean0.no/mean.no, type='l',
     xlab='Hazard Ratio', ylab='Recriprocal of Charge Ratio',
     xlim=c(0,3.5),ylim=c(0,10))
lines(elp, mean0.yes/mean.yes, lty=3)
abline(a=0,b=1,lty=2)
lines(elp, median0.no/median.no, lty=4)
lines(elp, median0.yes/median.yes, lty=5)
scat1d(exp(f$linear.predictors)) #show rug plot for density of exp(X*beta)
```


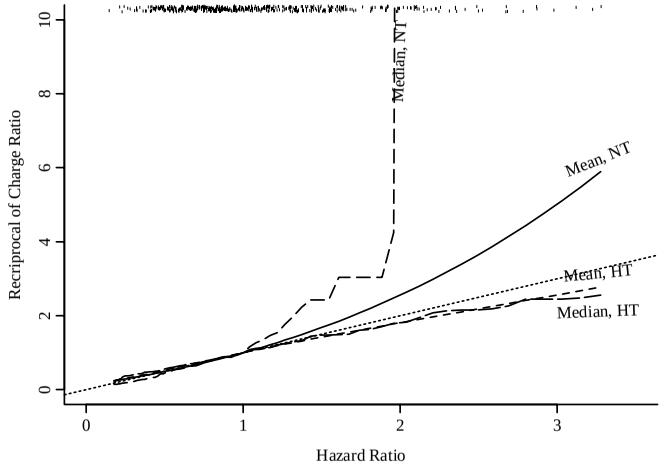

Figure 22: Relationships between hazard ratios and (mean or median) charge ratios

- Draw a nomogram depicting how predicted charges are computed

```
# Negate coefficients in model so that high risk = high charges
g ← f

g$coefficients ← -g$coef
g$center ← -g$center

mean.charge.no   ← function(lp) mean.charge(-lp, stratum=1)
mean.charge.yes  ← function(lp) mean.charge(-lp, stratum=2)
median.charge.no ← function(lp) quan(lp=-lp, stratum=1)
median.charge.yes← function(lp) quan(lp=-lp, stratum=2)

nomogram(g, dis=c(0,50,100),
         fun=list('Mean charge, NT'  =mean.charge.no,
                  'Mean charge, HT'  =mean.charge.yes,
                  'Median charge, NT'=median.charge.no,
                  'Median charge, HT'=median.charge.yes),
         lmgp=.4,lp.at=seq(-1.2,1.2,by=.2),
         fun.at=c(10,20,30,40,50,100,150,200,250,300,400,500,600,700,800))
```

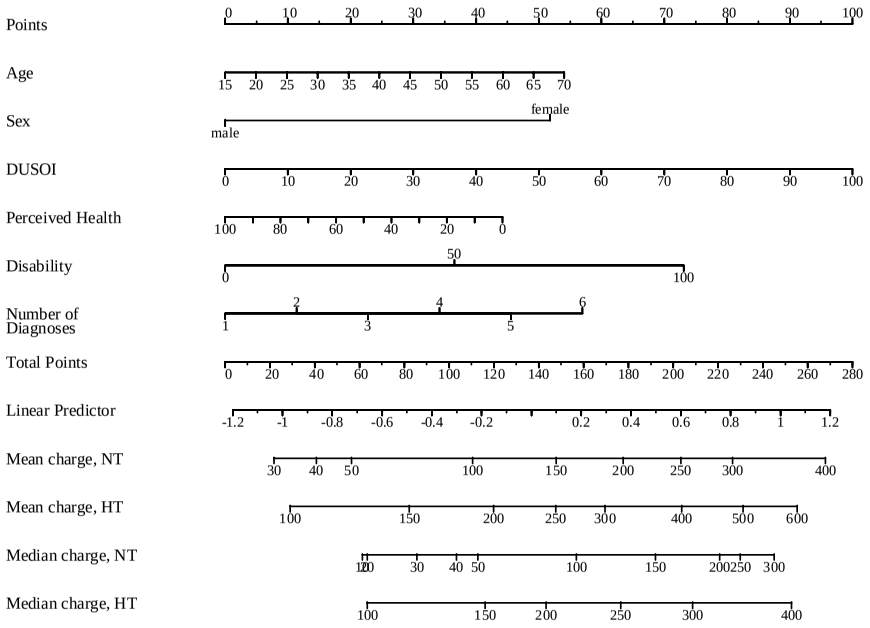

Figure 23: Nomogram depicted fitted stratified Cox PH model

References

[1] E. Arjas. A graphical method for assessing goodness of fit in Cox's proportional hazards model. *Journal of the American Statistical Association*, 83:204–212, 1988.

[2] A. C. Atkinson. A note on the generalized information criterion for choice of a model. *Biometrika*, 67:413–418, 1980.

[3] D. R. Cox. Regression models and life-tables (with discussion). *Journal of the Royal Statistical Society B*, 34:187–220, 1972.

[4] C. de Boor. *A Practical Guide to Splines*. Springer-Verlag, New York, 1978.

[5] T. F. Devlin and B. J. Weeks. Spline functions for logistic regression modeling. In *Proceedings of the Eleventh Annual SAS Users Group International Conference*, pages 646–651, Cary, NC, 1986. SAS Institute, Inc.

[6] R. Dudley, F. E. Harrell, L. Smith, D. B. Mark, R. M. Califf, D. B. Pryor, D. Glower, J. Lipscomb, and M. Hlatky. Comparison of analytic models for estimating the effect of clinical factors on the cost of coronary artery bypass graft surgery. *Journal of Clinical Epidemiology*, 46:261–271, 1993.

[7] S. Durrleman and R. Simon. Flexible regression models with cubic splines. *Statistics in Medicine*, 8:551–561, 1989.

[8] B. Efron. The efficiency of cox's likelihood function for censored data. *Journal of the American Statistical Association*, 72:557–565, 1977.

[9] J. J. Faraway. The cost of data analysis. *Journal of Computational and Graphical Statistics*, 1:213–229, 1992.

[10] P. Grambsch and T. Therneau. Proportional hazards tests and diagnostics based on weighted residuals. *Biometrika*, 81:515–526, 1994. Amendment and corrections in **82**, 668 (1995).

[11] F. E. Harrell, K. L. Lee, and D. B. Mark. Multivariable prognostic models: Issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. *Statistics in Medicine*, 14:to appear, 1995.

[12] F. E. Harrell, K. L. Lee, and B. G. Pollock. Regression models in clinical studies: determinging relationships between predictors and response. *Journal of the National Cancer Institute*, 80:1198–1202, 1988.

[13] K. R. Hess. Graphical methods for assessing violations of the proportional hazards assumption in Cox regression. *Statistics in Medicine*, 14:1707–1723, 1995.

[14] J. F. Lawless and C. Nadeau. Some simple robust methods for the analysis of recurrent events. *Technometrics*, 37:158–168, 1995.

[15] L. R. Muenz. Comparing survival distributions: A review for nonstatisticians. II. *Cancer Investigation*, 1:537–545, 1983.

[16] N. J. D. Nagelkerke. A note on a general definition of the coefficient of determination. *Biometrika*, 78:691–692, 1991.

[17] G. R. Parkerson, W. Broadhead, and C. J. Tse. Health status and severity of illness as predictors of outcomes in primary care. *Medical Care*, 33:53–66, 1995.

[18] D. Schoenfeld. Partial residuals for the proportional hazards regression model. *Biometrika*, 69:239–241, 1982.

[19] L. A. Sleeper and D. P. Harrington. Regression splines in the Cox model with application to covariate effects in liver disease. *Journal of the American Statistical Association*, 85:941–949, 1990.

[20] L. R. Smith, C. A. Milano, B. S. Molter, J. R. Elberry, D. C. Sabiston, and P. K. Smith. Preoperative determinants of postoperative costs associated with coronary artery bypass graft surgery. *Circulation*, 90 [part 2]:II–124–II–128, 1994.

[21] P. L. Smith. Splines as a useful and convenient statistical tool. *American Statistician*, 33:57–62, 1979.

[22] D. J. Spiegelhalter. Probabilistic prediction in patient management. *Statistics in Medicine*, 5:421–433, 1986.

[23] C. J. Stone. Comment: Generalized additive models. *Statistical Science*, 1:312–314, 1986.

[24] C. J. Stone and C. Y. Koo. Additive splines in statistics. In *Proceedings of the Statistical Computing Section ASA*, pages 45–48, 1985.

[25] T. M. Therneau, P. M. Grambsch, and T. R. Fleming. Martingale-based residuals for survival models. *Biometrika*, 77:216–218, 1990.

[26] J. C. van Houwelingen and S. le Cessie. Predictive value of statistical models. *Statistics in Medicine*, 8:1303–1325, 1990.

[27] P. Verweij and H. C. van Houwelingen. Penalized likelihood in Cox regression. *Statistics in Medicine*, 13:2427–2436, 1994.

This work was supported by grants from the Agency for Health Care Policy and Research (US Public Health Service) and the Robert Wood Johnson Foundation.

UNIX S–Plus functions are in the public domain in `statlib` (`Internet` address `lib.stat.cmu.edu`). To obtain the functions using E–mail, send the message `send Design from S` to `statlib@lib.stat.cmu.edu`.

To obtain a 550 page set of notes and S–PLUS function documentation, send a check for $38 to cover copying expenses made out to Duke University to:

Alicia McKinnis
Box 3363
Durham NC 27710