

---

## CASE STUDY IN PARAMETRIC SURVIVAL MODELING

Frank E Harrell Jr  
Department of Biostatistics  
Vanderbilt University School of Medicine  
Nashville TN 37232  
f.harrell@vanderbilt.edu  
biostat.mc.vanderbilt.edu/rms

---

ASA BIOPHARMACEUTICAL SECTION WEBINAR

3 APRIL 2009

## Course Philosophy

- Modeling and adjusting for covariables is necessary and informative in clinical trials
- Satisfaction of model assumptions improves precision and increases statistical power
- It is more productive to make a model fit step by step (e.g., transformation estimation) than to postulate a simple model and find out what went wrong
- Graphical methods should be married to formal inference
- Overfitting occurs frequently, so data reduction and model validation are important
- Carefully fitting an improper model is better than badly fitting (and overfitting) a well-chosen one
- Methods which work for all types of regression models are the most valuable.

- In most research projects the cost of data collection far outweighs the cost of data analysis, so it is important to use the most efficient and accurate modeling techniques, to avoid categorizing continuous variables, and to not remove data from the estimation sample just to be able to validate the model.
- The bootstrap is a breakthrough for statistical modeling and model validation.
- Using the data to guide the data analysis is almost as dangerous as not doing so.
- A good overall strategy is to decide how many degrees of freedom (i.e., number of regression parameters) can be “spent”, where they should be spent, to spend them with no regrets.

See the excellent text *Clinical Prediction Models* by Steyerberg<sup>42</sup>.

# **Covariable Adjustment in Randomized Clinical Trials**

# Covariable Adjustment in Linear Models

- Model:  $E(Y|X) = X\beta + \epsilon$
- Continuous response variable  $Y$ , normal residuals
- Statistical testing for baseline differences is scientifically incorrect (Altman & Doré 1990, Begg 1990, Senn 1994)
- If we are worried about baseline imbalance we need to search patient records for counter–balancing factors
- → imbalance is not the reason to adjust for covariables
- Adjust to gain efficiency by subtracting explained variation
- Relative efficiency of unadjusted treatment comparison is  $1 - \rho^2$
- Unadjusted analyses yields unbiased treatment effect estimate

# Covariable Adjustment in Nonlinear Models

- Gail, Wieand, Piantadosi 1984 showed that for unadjusted treatment estimates to be unbiased, regression must be linear or exponential
- Gail 1986 showed that for logistic, Cox, and paired survival models unadjusted treatment effects are asymptotically biased low in absolute value
- Gail also studied normal, exponential, additive risk, Poisson
- Senn [40, p. 3747] summarizes the problem as follows:

”Part of the problem with Poisson, proportional hazard and logistic regression approaches is that they use a single parameter, the linear predictor, with no equivalent of the variance parameter in the Normal case. This means that lack of fit impacts on the estimate of the predictor.”

# Cox / Log-Rank Test for Time to Event

- Lagakos & Schoenfeld 1984 showed that type I error is preserved if don't adjust
- If hazards are proportional conditional on covariables, they are not proportional if omit covariables
- Morgan 1986 derived asymptotic relative efficiencies (ARE) of unadjusted log-rank test if a binary covariable is omitted
- If prevalence of covariable  $X$  is 0.5:

## Efficiency of Unadjusted Log-Rank Test

| $X = 1 : X = 0$ Hazard Ratio | ARE  |
|------------------------------|------|
| 1.0                          | 1.00 |
| 1.5                          | 0.95 |
| 2.0                          | 0.88 |
| 3.0                          | 0.72 |

- Ford, Norrie, Ahmadi 1995: Treatment effect does not have the same interpretation under unadjusted and adjusted models
- No reason for the two hazard ratios to have the same value

Akazawa, Nakamura, Palesch 1997: Power of unadjusted and stratified log-rank test

| Number of Strata | Range of Log Hazards | Power  |          |
|------------------|----------------------|--------|----------|
|                  |                      | Unadj. | Adjusted |
| 1                | 0                    | .78    | –        |
| 2                | 0–0.5                | .77    | .78      |
|                  | 0–1                  | .67    | .78      |
|                  | 0–2                  | .36    | .77      |
| 4                | 0–3                  | .35    | .77      |
| 8                | 0–3.5                | .33    | .77      |

## Sample Size Calculation Issues

- Schoenfeld 1983 implies that covariable adjustment can only ↑ sample size in randomized trials
- Need to recognize ill–definition of unadjusted hazard ratios

# Why are Adjusted Estimates Right?

- Hauck, Anderson, Marcus<sup>17</sup>, who have an excellent review of covariable adjustment in nonlinear models, state:

“For use in a clinician–patient context, there is only a single person, that patient, of interest. The subject-specific measure then best reflects the risks or benefits for that patient. Gail has noted this previously [ENAR Presidential Invited Address, April 1990], arguing that one goal of a clinical trial ought to be to predict the direction and size of a treatment benefit for a patient with specific covariate values. In contrast, population–averaged estimates of treatment effect compare outcomes in groups of patients. The groups being compared are determined by whatever covariates are included in the model. The treatment effect is then a comparison of average outcomes, where the averaging is over all omitted covariates.”

# How Many Covariables to Use?

- Try to adjust for the bulk of the variation in outcome<sup>17, 44</sup>
- Neuhaus<sup>31</sup>: “to improve the efficiency of estimated covariate effects of interest, analysts of randomized clinical trial data should adjust for covariates that are strongly associated with the outcome”
- Raab *et al.*<sup>34</sup> have more guidance for choosing covariables and provide a formula for linear model that shows how the value of adding a covariable depends on the sample size

# Statistical Plan for Randomized Trials

- When a relevant dataset is available before the trial begins, develop the model from the dataset and use the predicted value as a single adjustment covariable in the trial (Knaus et al. 1993)
- Otherwise: CPMP Working Party: Finalize choice of model, transformations, interactions before merging treatment assignment into analysis dataset.  
Edwards<sup>9</sup>: Pre-specify family of models that will be used, along with the strategy for selecting the particular model.  
Masked model derivation does not bias treatment effect.
- CPMP guidance<sup>5</sup>
  - “Stratification may be used to ensure balance of treatments across covariates; it may also be used for administrative reasons. The factors that are the basis of stratification should normally be included as covariates in the primary model.
  - Variables known a priori to be strongly, or at least moderately, associated with the primary outcome and/or variables

for which there is a strong clinical rationale for such an association should also be considered as covariates in the primary analysis. The variables selected on this basis should be pre-specified in the protocol or the statistical analysis plan.

- Baseline imbalance observed post hoc should not be considered an appropriate reason for including a variable as a covariate in the primary analysis.
- Variables measured after randomization and so potentially affected by the treatment should not normally be included as covariates in the primary analysis.
- If a baseline value of a continuous outcome measure is available, then this should usually be included as a covariate. This applies whether the primary outcome variable is defined as the 'raw outcome' or as the 'change from baseline'.
- Only a few covariates should be included in a primary analysis. Although larger data sets may support more covariates than smaller ones, justification for including each of the co-

- variates should be provided. (???)
- In the absence of prior knowledge, a simple functional form (usually either linearity or dichotomising a continuous scale) should be assumed for the relationship between a continuous covariate and the outcome variable. (???)
  - The validity of the model assumptions must be checked when assessing the results. This is particularly important for generalized linear or non-linear models where mis-specification could lead to incorrect estimates of the treatment effect. Even under ordinary linear models, some attention should be paid to the possible influence of extreme outlying values.
  - Whenever adjusted analyses are presented, results of the treatment effect in subgroups formed by the covariates (appropriately categorised, if relevant) should be presented to enable an assessment of the validity of the model assumptions. (???)
  - Sensitivity analyses should be pre-planned and presented to investigate the robustness of the primary results. Discrepancies should be discussed and explained. In the pres-

ence of important differences that cannot be logically explained- for example, between the results of adjusted and unadjusted analyses-the interpretation of the trial could be seriously affected.

- The primary model should not include treatment by covariate interactions. If substantial interactions are expected a priori, the trial should be designed to allow separate estimates of the treatment effects in specific subgroups.
- Exploratory analyses may be carried out to improve the understanding of covariates not included in the primary analysis, and to help the sponsor with the ongoing development of the drug.
- A primary analysis, unambiguously pre-specified in the protocol or statistical analysis plan, correctly carried out and interpreted, should support the conclusions which are drawn from the trial. Since there may be a number of alternative valid analyses, results based on pre-specified analyses will carry most credibility.”

“In confirmatory trials, a model is pre-specified, and it is necessary to pretend that it is true. In most other statistical applications, the choice of model is data-driven, but it is necessary to pretend that it is not.”<sup>9</sup>

See also Siqueira and Taylor<sup>41</sup>.

- Choose predictors based on expert opinion
- Impute missing values rather than discarding observations
- Keep all pre-specified predictors in model, regardless of  $P$ -value
- Use shrinkage (penalized maximum likelihood estimation) to avoid over-adjustment
- Detailed strategy in REGRESSION MODELING STRATEGIES, Springer, 2001.

## Covariable Adjustment vs. Allocation Based on Covariates

As Senn<sup>40</sup> states,

"The decision to fit prognostic factors has a far more dramatic effect on the precision of our inferences than the choice of an allocation based on covariates or randomization approach and one of my chief objections to the allocation based on covariates approach is that trialists have tended to use the fact that they have balanced as an excuse for not fitting. This is a grave mistake." (p. 3748)

"My view . . . was that the form of analysis envisaged (that is to say, which factors and covariates should be fitted) justified the allocation and *not vice versa*." (p. 3747)

# Summary

As Senn [40, p. 3741] said

”The point of view is sometimes defended that analyses that ignore covariates are superior because they are simpler. I do not accept this. A value of  $\pi = 3$  is a simple one and accurate to one significant figure . . . However very few would seriously maintain that it should generally be adopted by engineers.”

# Notes

I think it is most important to decide what it is you want to estimate, and then formulate a model that will accomplish that. Unlike ordinary linear models, which provide unbiased treatment effects if balanced covariates are mistakenly omitted from the model in an RCT, most models (such as the Cox PH model) result in biased treatment effects even when there is perfect balance in covariates, if the covariates have nonzero effects on the outcome. This is another way of talking about residual outcome heterogeneity.

If you want to estimate the effect of variable  $X$  on survival time, averaging over males and females in some strange undocumented way, you can get the population averaged effect of  $X$  without including sex in the model. Recognize however this is like comparing some of the males with some of the females when estimating the  $X$  effect. This is seldom of interest. More likely we want to know the effect of  $X$  for males, the effect for females, and if there is no interaction we pool the two to more precisely estimate the effect of  $X$  conditional on sex.

Another way to view this is that the PH assumption is more likely to hold when you condition on covariates than when you don't. No matter what happens though, if PH holds for one case, it cannot hold for the other, e.g., if PH holds after conditioning, it cannot hold when just looking at the marginal effect of  $X$ .

# Parametric Survival Models

# Homogeneous Models (No Predictors)

Why use a parametric model?

1. easily compute selected quantiles of the survival distribution
2. estimate (usually by extrapolation) the expected failure time
3. derive a concise equation and smooth function for estimating  $S(t)$ ,  $\Lambda(t)$ , and  $\lambda(t)$
4. estimate  $S(t)$  more precisely than  $S_{KM}(t)$  or  $S_{\Lambda}(t)$  (Altschuler-Nelson-Fleming-Harrington estimator) if the parametric form is correctly specified.

**Note:** Fitting more than two smooth survival curves and choosing the one that best reproduces the KM estimator will result in a true precision no better than KM.

## Specific Models

- Seen exponential and Weibull already
- Many others obtained by assuming  $\log(T)$  has a certain dist.
- Log-normal:  $S(t) = 1 - \Phi\left(\frac{\log(t)-\mu}{\sigma}\right)$
- Log-logistic:  $S(t) = [1 + \exp(-\frac{\log(t)-\mu}{\sigma})]^{-1}$
- Log-extreme value:  $S(t) = \exp[-\exp(\frac{\log(t)-\mu}{\sigma})]$   
another way of expressing Weibull

# Estimation

- Log-likelihood for exponential distribution

$$\log L = \sum_{i: Y_i \text{ uncensored}}^n \log \lambda - \sum_{i=1}^n \lambda Y_i.$$

$$\hat{\lambda} = n_u/w$$

$$\text{var}(\hat{\lambda}) = n_u/w^2$$

$$\text{var}(\log \hat{\lambda}) = 1/n_u$$

$$\hat{\mu} = w/n_u$$

$$\hat{S}(t) = \exp(-\hat{\lambda}t)$$

1 3 3 6<sup>+</sup> 8<sup>+</sup> 9 10<sup>+</sup>.

$$n_u = 4$$

$$\begin{aligned}w &= 40 \\ \hat{\mu} &= 10 \pm 5 \\ T_{0.5} &= 10 \log(2)\end{aligned}$$

- Weibull fit

$$\begin{aligned}\hat{\alpha} &= 0.0728 \\ \hat{\gamma} &= 1.164 \\ \hat{S}(t) &= \exp(-0.0728t^{1.164}) \\ \hat{S}^{-1}(0.5) &= [(\log 2)/\hat{\alpha}]^{1/\hat{\gamma}} = 6.935 \\ &\text{(estimated median).}\end{aligned}$$

## Assessment of Model Fit

- Example: Weibull

$$\log[-\log S(t)] = \log \Lambda(t) = \log \alpha + \gamma(\log t).$$

- Plot  $\log \hat{\Lambda}(t)$  versus  $\log t$
- For assumed dist.  $S(t)$  plot  $S^{-1}[S_{\Lambda}(t)]$  or  $S^{-1}[S_{KM}(t)]$  against  $t$ , check for linearity
- Log-distributions: plot vs.  $\log t$
- Check log-normal: plot  $\Phi^{-1}[S_{\Lambda}(t)]$  vs.  $\log t$
- Check log-logistic: plot  $\text{logit}[S_{\Lambda}(t)]$  vs.  $\log t$
- Alternative: plot fitted  $\hat{S}(t)$  and  $S_{\Lambda}(t)$  vs.  $t$  on the same graph

# Parametric Proportional Hazards Models

## Model

$$\lambda(t|X) = \lambda(t) \exp(X\beta).$$

$$\Lambda(t|X) = \Lambda(t) \exp(X\beta)$$

$$S(t|X) = \exp[-\Lambda(t) \exp(X\beta)] = \exp[-\Lambda(t)]^{\exp(X\beta)}.$$

$$S(t|X) = S(t)^{\exp(X\beta)},$$

## Model Assumptions and Interpretation of Parameters

$$\begin{aligned}\log \lambda(t|X) &= \log \lambda(t) + X\beta \\ \log \Lambda(t|X) &= \log \Lambda(t) + X\beta.\end{aligned}$$

Assumptions:

- Underlying functions ( $\lambda$ ,  $\Lambda$ ,  $S$ )
- Linear effect of predictors on  $\log \lambda$ ,  $\log \Lambda$
- No interaction between  $X$  and  $t \rightarrow$  impact same over time

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

- Effect of increasing  $X_j$  by  $d$  is to increase  $\lambda$  by factor of  $\exp(\beta_j d)$

- One binary predictor:

$$\lambda(t|X_1 = 0) = \lambda(t)$$

$$\lambda(t|X_1 = 1) = \lambda(t) \exp(\beta_1).$$

Here  $\exp(\beta_1)$  is the  $X_1 = 1 : X_1 = 0$  hazard ratio.

- One continuous predictor:

$$\lambda(t|X_1) = \lambda(t) \exp(\beta_1 X).$$

## Hazard Ratio, Risk Ratio, and Risk Difference

$$S_T = S_C^{0.5}$$

| Subject | 5-Year Survival |      | Difference | Mortality Ratio (T/C) |
|---------|-----------------|------|------------|-----------------------|
|         | C               | T    |            |                       |
| A       | 0.98            | 0.99 | 0.01       | $0.01/0.02 = 0.5$     |
| B       | 0.80            | 0.89 | 0.09       | $0.11/0.2 = 0.55$     |
| C       | 0.25            | 0.50 | 0.25       | $0.5/0.75 = 0.67$     |

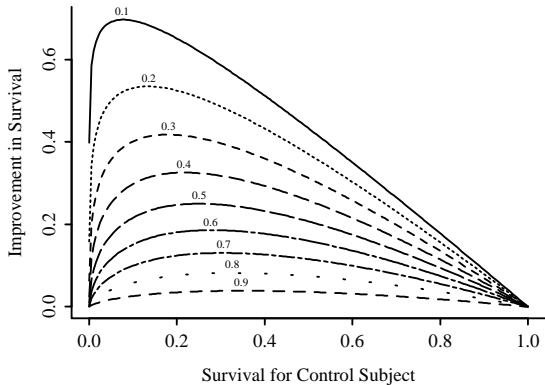


Figure 1: Absolute clinical benefit as a function of survival in a control subject and the relative benefit (hazard ratio). The hazard ratios are given for each curve.

## Specific Models

- Exponential:

$$\lambda(t|X) = \lambda \exp(X\beta)$$

$$S(t|X) = \exp[-\lambda t \exp(X\beta)] = \exp(-\lambda t)^{\exp(X\beta)}.$$

$$E\{T|X\} = 1/[\lambda \exp(X\beta)]$$

$$T_{0.5}|X = (\log 2)/[\lambda \exp(X\beta)].$$

- Weibull:

$$\lambda(t|X) = \alpha\gamma t^{\gamma-1} \exp(X\beta)$$

$$\Lambda(t|X) = \alpha t^\gamma \exp(X\beta)$$

$$\begin{aligned} S(t|X) &= \exp[-\alpha t^\gamma \exp(X\beta)] \\ &= [\exp(-\alpha t^\gamma)]^{\exp(X\beta)}. \end{aligned}$$

$$T_{0.5}|X = \{\log 2/[\alpha \exp(X\beta)]\}^{1/\gamma}.$$

For numerical reasons, re-write:

$$\begin{aligned} S(t|X) &= \exp(-\Lambda(t|X)), \quad \text{where} \\ \Lambda(t|X) &= \exp(\gamma \log t + X\beta). \end{aligned}$$

See also spline hazard models<sup>20, 21, 26</sup> and the generalized gamma distribution<sup>6</sup>.

## Assessment of Model Fit

If  $\lambda(t)$  is Weibull, the two curves will be linear if  $\log t$  is plotted instead of  $t$  on the  $x$ -axis.

- Weibull: Stratify on  $X$ , plot  $\log \Lambda_{KM}(t|X \text{ stratum})$  vs.  $\log t$ .
- Assesses PH in addition to shape assumptions—all curves should be parallel as well as straight.

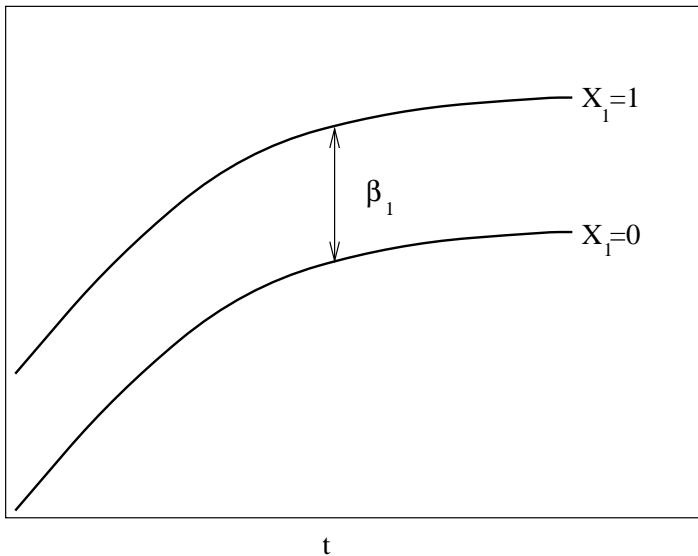
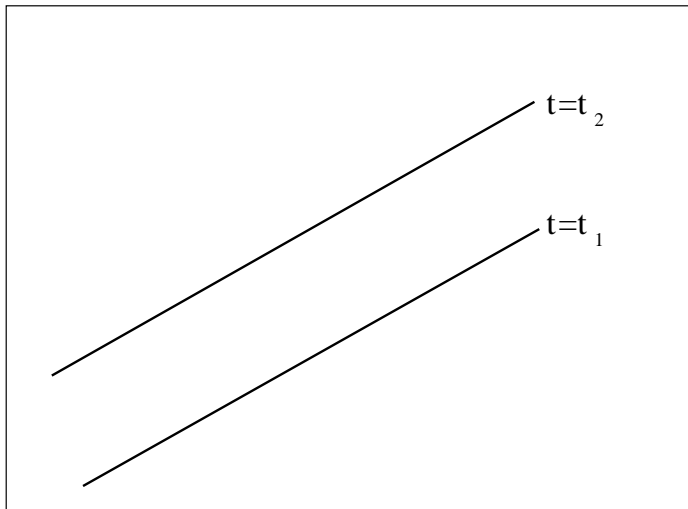


Figure 2: PH Model with one binary predictor. Y-axis is  $\log \lambda(t)$  or  $\log \Lambda(t)$ . For  $\log \Lambda(t)$ , the curves must be non-decreasing. For  $\log \lambda(t)$ , they may be any shape.



$X_1$

**Figure 3:** PH model with one continuous predictor. Y-axis is  $\log \lambda(t)$  or  $\log \Lambda(t)$ . For  $\log \Lambda(t)$ , drawn for  $t_2 > t_1$ . The slope of each line is  $\beta_1$ .

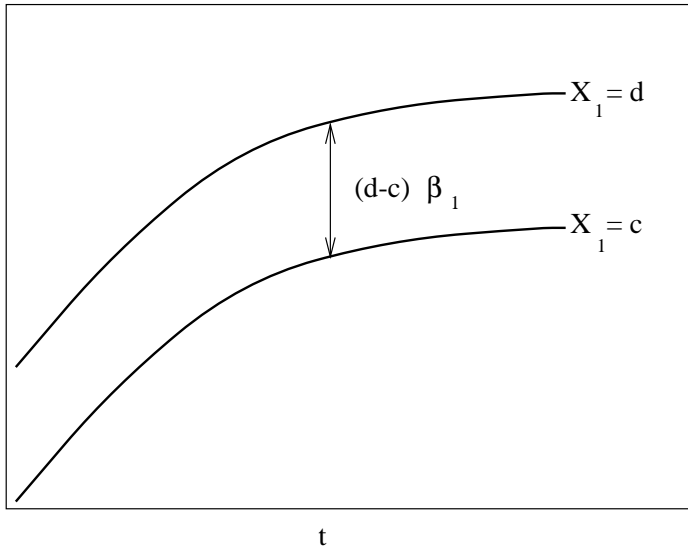


Figure 4: PH model with one continuous predictor. Y-axis is  $\log \lambda(t)$  or  $\log \Lambda(t)$ . For  $\log \lambda$ , the functions need not be monotonic.

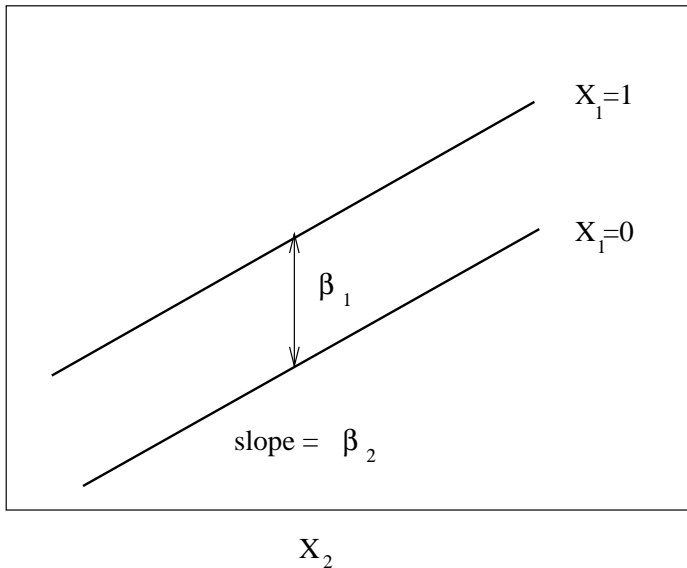


Figure 5: Regression assumptions, linear additive PH or AFT model with two predictors. For PH,  $Y$ -axis is  $\log \lambda(t)$  or  $\log \Lambda(t)$  for a fixed  $t$ . For AFT,  $Y$ -axis is  $\log(T)$ .

# Accelerated Failure Time Models

## Model

- Specifies that predictors act multiplicatively on failure time
- Alters rate subject proceeds along time axis

$$S(t|X) = \psi\left(\frac{\log(t) - X\beta}{\sigma}\right),$$

$$\frac{\log(T) - X\beta}{\sigma} \sim \psi$$

$$\log(T) = X\beta + \sigma\epsilon$$

$$\epsilon \sim \psi$$

- Weibull (and exponential) members of PH and AFT

## Model Assumptions and Interpretation of Parameters

$$\psi^{-1}[S(t|X)] = \frac{\log(t) - X\beta}{\sigma}.$$

Letting  $\epsilon \sim \psi$

$$\log(T) = X\beta + \sigma\epsilon.$$

Check that residuals  $\log(T) - X\hat{\beta} \sim \psi$  (within scale factor).

The assumptions of the AFT model are thus

1. The true form of  $\psi$  (the distributional family) is correctly specified.
2. In the absence of nonlinear and interaction terms, each  $X_j$  affects  $\log(T)$  or  $\psi^{-1}[S(t|X)]$  linearly.
3. Implicit in these assumptions is that  $\sigma$  is a constant independent of  $X$ .

1-unit change in  $X_j = \beta_j$  change in  $\log T$ , or increase  $T$  by factor of  $\exp(\beta_j)$ .

Median survival time:

$$T_{0.5}|X = \exp[X\beta + \sigma\psi^{-1}(0.5)]$$

## Specific Models

- Extreme value:  $\psi(u) = \exp(-\exp(u))$
- Logistic:  $\psi(u) = [1 + \exp(u)]^{-1}$
- Normal:  $\psi(u) = 1 - \Phi(u)$
- Log-normal:

$$S(t|X) = 1 - \Phi\left(\frac{\log(t) - X\beta}{\sigma}\right),$$

- Log-logistic:

$$S(t|X) = [1 + \exp\left(\frac{\log(t) - X\beta}{\sigma}\right)]^{-1}.$$

## Estimation

Works better if  $\sigma$  parameterized as  $\exp(\delta)$ .

$$\begin{aligned}\hat{S}(t|X) &= \psi\left(\frac{\log(t) - X\hat{\beta}}{\hat{\sigma}}\right) \\ \hat{T}_{0.5}|X &= \exp[X\hat{\beta} + \hat{\sigma}\psi^{-1}(0.5)].\end{aligned}$$

Normal and logistic:  $\hat{T}_{0.5}|X = \exp(X\hat{\beta})$ .

$$\psi\left(\frac{\log(t) - X\hat{\beta}}{\hat{\sigma}} \pm z_{1-\alpha/2} \times s\right).$$

## Residuals

For an AFT model, standardized residuals are simply

$$r = (\log(T) - X\hat{\beta})/\sigma.$$

When  $T$  is right-censored,  $r$  is right-censored.

## Assessment of Model Fit

|         |                  |     |     |     |     |     |     |                  |                  |                  |
|---------|------------------|-----|-----|-----|-----|-----|-----|------------------|------------------|------------------|
| Group 1 | 143              | 164 | 188 | 188 | 190 | 192 | 206 | 209              | 213              | 216              |
|         | 220              | 227 | 230 | 234 | 246 | 265 | 304 | 216 <sup>+</sup> | 244 <sup>+</sup> |                  |
| Group 2 | 142              | 156 | 163 | 198 | 205 | 232 | 232 | 233              | 233              | 233              |
|         | 233              | 239 | 240 | 261 | 280 | 280 | 296 | 296              | 323              | 204 <sup>+</sup> |
|         | 344 <sup>+</sup> |     |     |     |     |     |     |                  |                  |                  |

$$S_{\text{extreme}}(t) = \exp\left[-\exp\left(\frac{\log(t) - 5.45 - 0.132\{\text{Group 2}\}}{0.183}\right)\right]$$

$$S_{\text{log-logistic}}(t) = \left[1 + \exp\left(\frac{\log(t) - 5.38 - 0.105\{\text{Group 2}\}}{0.116}\right)\right]^{-1}$$

$$S_{\text{log-normal}}(t) = 1 - \Phi\left(\frac{\log(t) - 5.38 - 0.0931\{\text{Group 2}\}}{0.210}\right)$$

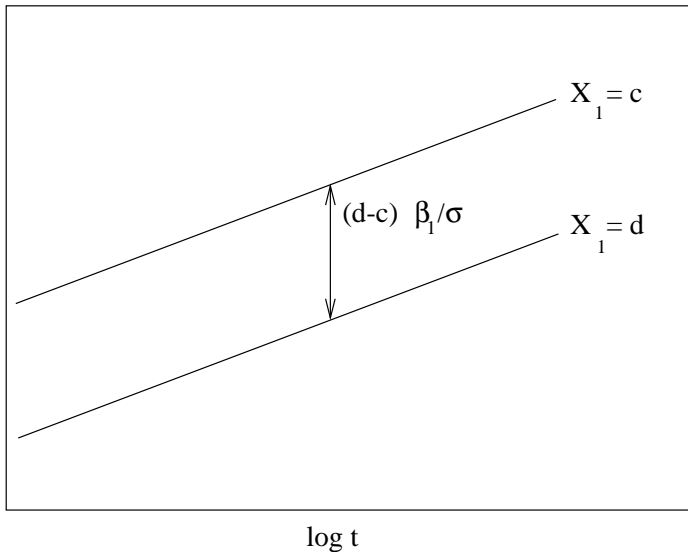
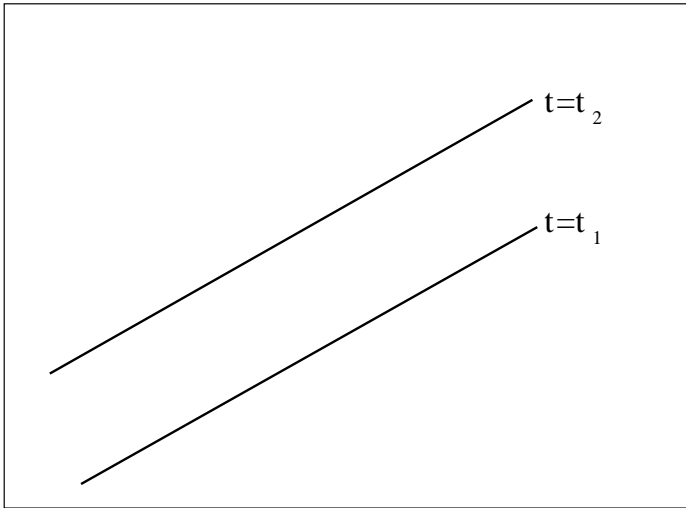


Figure 6: AFT model with one predictor. Y-axis is  $\psi^{-1}[S(t|X)] = \frac{\log(t) - X\beta}{\sigma}$ . Drawn for  $d > c$ . The slope of the lines is  $\sigma^{-1}$ .



$X_1$

Figure 7: AFT model with one continuous predictor. Y-axis is  $\psi^{-1}[S(t|X)] = \frac{\log(t) - X\beta}{\sigma}$ . Drawn for  $t_2 > t_1$ . The slope of each line is  $\beta_1/\sigma$  and the difference between the lines is  $\frac{1}{\sigma} \log(t_2/t_1)$ .

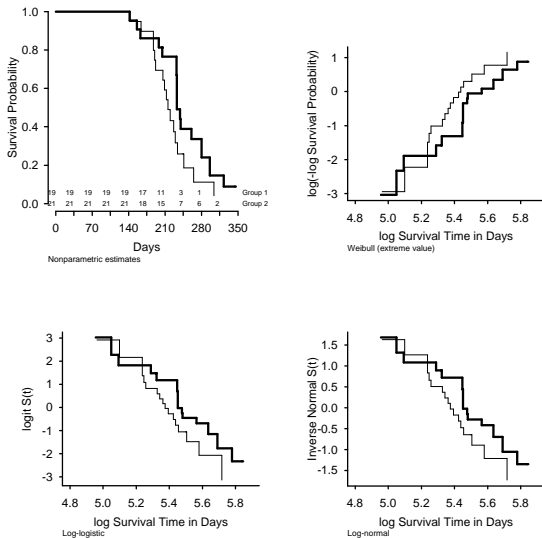


Figure 8: Altschuler-Nelson-Fleming-Harrington nonparametric survival estimates for rats treated with DMBA<sup>32</sup>, along with various transformations of the estimates for checking distributional assumptions of 3 parametric survival models.

| Model                   | Group 2:1<br>Failure Time Ratio | Median Survival Time |         |
|-------------------------|---------------------------------|----------------------|---------|
|                         |                                 | Group 1              | Group 2 |
| Extreme Value (Weibull) | 1.14                            | 217                  | 248     |
| Log-logistic            | 1.11                            | 217                  | 241     |
| Log-normal              | 1.10                            | 217                  | 238     |

### Weibull PH form

$$S(t|X) = \exp[-\exp(5.46 \log(t) - 29.781 - 0.721\{\text{Group } 2\})]$$

- More general approach to verifying distributional assumptions:
- Plot nonparametric estimate of survival distribution of  $r$
- Superimpose theoretical standardized distribution
- Can get distribution of residuals separately by strata — should all have same standardized distribution (e.g., same  $\sigma$ )

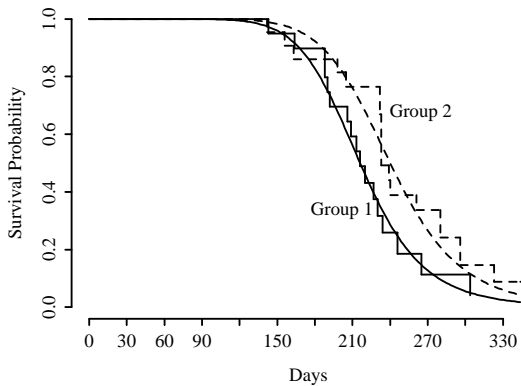


Figure 9: Agreement between fitted log-logistic model and nonparametric survival estimates for rat vaginal cancer data

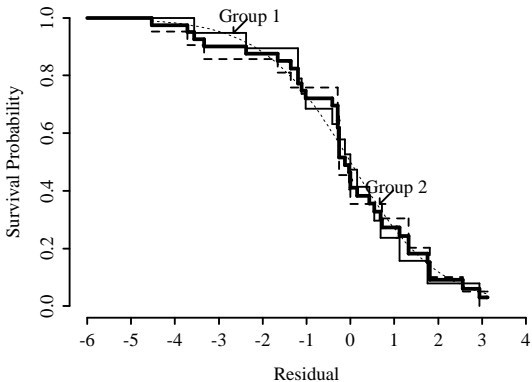


Figure 10: Kaplan-Meier estimates of distribution of standardized, censored residuals from the log-logistic model, along with the assumed standard log-logistic distribution (dashed curve). Thick step function is the estimated distribution of all residuals, and other step functions are the estimated distributions of residuals stratified by group, as indicated.

## Validating the Fitted Model

- Check distributional shape
- Group predicted  $t$ -year survival and plot Kaplan-Meier estimate at  $t$  vs. mean predicted  $\hat{S}$
- Cox-Snell residuals — check against  $U[0, 1]$
- `loess` smooth of  $F(T|X) - 0.5F(C|X)$  against  $X\hat{\beta}$  or  $\frac{2F(T|X)}{F(C|X)}$  vs.  $X\hat{\beta}$  if  $C$  is known

See the `val.surv` function in the `Design` package.

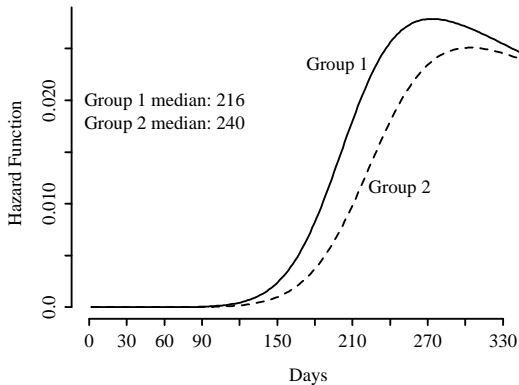


Figure 11: Estimated hazard functions for log-logistic fit to rat vaginal cancer data, along with median survival times

# **Case Study in Parametric Survival Modeling and Model Approximation**

**Data source:** Random sample of 1000 patients from Phases I & II of SUPPORT (Study to Understand Prognoses Preferences Outcomes and Risks of Treatment, funded by the Robert Wood Johnson Foundation). See <sup>25</sup>. The dataset is available from <http://biostat.mc.vanderbilt.edu/DataSets>.

- Analyze acute disease subset of SUPPORT (acute respiratory failure, multiple organ system failure, coma) — the shape of the survival curves is different between acute and chronic disease categories
- Patients had to survive until day 3 of the study to qualify
- Baseline physiologic variables measured during day 3

# Descriptive Statistics

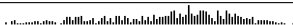
Create a variable `acute` to flag categories of interest; print univariable descriptive statistics.

```
> library(Hmisc,T); library(Design,T)
> acute ← support$dzclass %in% c('ARF/MOSF','Coma')
> describe(support[acute,]) # actually used latex(describe)
```

support[acute, ]  
35 Variables      537 Observations

---

age : Age



| n   | missing | unique | Mean | .05   | .10   | .25   | .50   | .75   | .90   | .95   |
|-----|---------|--------|------|-------|-------|-------|-------|-------|-------|-------|
| 537 | 0       | 529    | 60.7 | 28.49 | 35.22 | 47.93 | 63.67 | 74.49 | 81.54 | 85.56 |

lowest : 18.04 18.41 19.76 20.30 20.31  
highest: 91.62 91.82 91.93 92.74 95.51

---

**death : Death at any time up to NDI date:31DEC94**

|  | n   | missing | unique | Sum | Mean   |
|--|-----|---------|--------|-----|--------|
|  | 537 | 0       | 2      | 356 | 0.6629 |

---

**sex**

|  | n   | missing | unique |
|--|-----|---------|--------|
|  | 537 | 0       | 2      |

female (251, 47%), male (286, 53%)

---

**hospdead : Death in Hospital**

|  | n   | missing | unique | Sum | Mean   |
|--|-----|---------|--------|-----|--------|
|  | 537 | 0       | 2      | 201 | 0.3743 |

---

**slos : Days from Study Entry to Discharge**



|  | n   | missing | unique | Mean  | .05 | .10 | .25 | .50  | .75  | .90  | .95  |
|--|-----|---------|--------|-------|-----|-----|-----|------|------|------|------|
|  | 537 | 0       | 85     | 23.44 | 4.0 | 5.0 | 9.0 | 15.0 | 27.0 | 47.4 | 68.2 |

lowest : 3 4 5 6 7, highest: 145 164 202 236 241

---

**d.time : Days of Follow-Up**

|  | n   | missing | unique | Mean  | .05 | .10 | .25 | .50 | .75 | .90  | .95  |
|--|-----|---------|--------|-------|-----|-----|-----|-----|-----|------|------|
|  | 537 | 0       | 340    | 446.1 | 4   | 6   | 16  | 182 | 724 | 1421 | 1742 |

lowest : 3 4 5 6 7, highest: 1977 1979 1982 2011 2022

---

**dzgroup : Disease Group**

|  | n   | missing | unique |
|--|-----|---------|--------|
|  | 537 | 0       | 3      |

ARF/MOSF w/Sepsis (391, 73%), Coma (60, 11%), MOSF w/Malig (86, 16%)

---

**dzclass : Disease Class**

|  | n   | missing | unique |
|--|-----|---------|--------|
|  | 537 | 0       | 2      |

ARF/MOSF (477, 89%), Coma (60, 11%)

---

**num.co : number of comorbidities**

```
      n missing unique  Mean
537 0          7      1.525
Frequency  0   1   2   3   4   5   6
           111 196 133 51 31 10  5
           %  21  36  25  9  6  2  1
```

---

**edu : Years of Education**

```
      n missing unique  Mean .05 .10 .25 .50 .75 .90 .95
411 126      22      12.03 7  8 10 12 14 16 17
lowest : 0 1 2 3 4, highest: 17 18 19 20 22
```

---

**income**

```
      n missing unique
335      202      4
under $11k (158, 47%), $11-$25k (79, 24%), $25-$50k (63, 19%)
>$50k (35, 10%)
```

---

### scoma : SUPPORT Coma Score based on Glasgow D3

|           | n   | missing | unique | Mean  | .05 | .10 | .25 | .50 | .75 | .90 | .95 |     |   |
|-----------|-----|---------|--------|-------|-----|-----|-----|-----|-----|-----|-----|-----|---|
|           | 537 | 0       | 11     | 19.24 | 0   | 0   | 0   | 0   | 37  | 55  | 100 |     |   |
| Frequency |     | 0       | 9      | 26    | 37  | 41  | 44  | 55  | 61  | 89  | 94  | 100 |   |
| %         |     | 301     | 50     | 44    | 19  | 17  | 43  | 11  | 6   | 8   | 6   | 32  |   |
|           |     | %       | 56     | 9     | 8   | 4   | 3   | 8   | 2   | 1   | 1   | 1   | 6 |

---

### charges : Hospital Charges



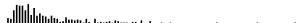
|          | n   | missing | unique | Mean   | .05    | .10    | .25   | .50   | .75    | .90    | .95    |
|----------|-----|---------|--------|--------|--------|--------|-------|-------|--------|--------|--------|
|          | 517 | 20      | 516    | 86652  | 11075  | 15180  | 27389 | 51079 | 100904 | 205562 | 283411 |
| lowest : |     | 3448    | 4432   | 4574   | 5555   | 5849   |       |       |        |        |        |
| highest: |     | 504660  | 538323 | 543761 | 706577 | 740010 |       |       |        |        |        |

---

### totcst : Total RCC cost



|          | n   | missing | unique | Mean   | .05    | .10    | .25   | .50   | .75   | .90    | .95    |
|----------|-----|---------|--------|--------|--------|--------|-------|-------|-------|--------|--------|
|          | 471 | 66      | 471    | 46360  | 6359   | 8449   | 15412 | 29308 | 57028 | 108927 | 141569 |
| lowest : |     | 0       | 2071   | 2522   | 3191   | 3325   |       |       |       |        |        |
| highest: |     | 269057  | 269131 | 338955 | 357919 | 390460 |       |       |       |        |        |

**totmst : Total micro-cost**

|          | n   | missing | unique | Mean  | .05  | .10  | .25   | .50   | .75   | .90   | .95    |
|----------|-----|---------|--------|-------|------|------|-------|-------|-------|-------|--------|
| lowest : | 331 | 206     | 328    | 39022 | 6131 | 8283 | 14415 | 26323 | 54102 | 87495 | 111920 |
| highest: | 0   | 1562    | 2478   | 2626  | 3421 |      |       |       |       |       |        |

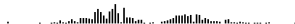
**avtisst : Average TISS, Days 3-25**

|          | n     | missing | unique | Mean  | .05   | .10   | .25   | .50   | .75   | .90   | .95   |
|----------|-------|---------|--------|-------|-------|-------|-------|-------|-------|-------|-------|
| lowest : | 536   | 1       | 205    | 29.83 | 12.46 | 14.50 | 19.62 | 28.00 | 39.00 | 47.17 | 50.37 |
| highest: | 4.000 | 5.667   | 8.000  | 9.000 | 9.500 |       |       |       |       |       |       |

**race : Race**

|           | n   | missing | unique |       |       |       |       |          |
|-----------|-----|---------|--------|-------|-------|-------|-------|----------|
| Frequency | 535 | 2       | 5      | white | black | asian | other | hispanic |
| %         |     |         |        | 78    | 16    | 1     | 1     | 4        |

### meanbp : Mean Arterial Blood Pressure Day 3



```
      n missing unique Mean .05 .10 .25 .50 .75 .90 .95
537 0          109    83.28 41.8 49.0 59.0 73.0 111.0 124.4 135.0
lowest : 0 20 27 30 32, highest: 155 158 161 162 180
```

---

### wbc : White Blood Cell Count Day 3



```
      n missing unique Mean .05 .10 .25 .50 .75 .90
532 5          241    14.1 0.8999 4.5000 7.9749 12.3984 18.1992 25.1891
      .95
30.1873
lowest : 0.05000 0.06999 0.09999 0.14999 0.19998
highest: 51.39844 58.19531 61.19531 79.39062 100.00000
```

---

### hrt : Heart Rate Day 3



```
      n missing unique Mean .05 .10 .25 .50 .75 .90 .95
537 0          111    105 51 60 75 111 126 140 155
lowest : 0 11 30 36 40, highest: 189 193 199 232 300
```

---

### resp : Respiration Rate Day 3



```
      n missing unique Mean .05 .10 .25 .50 .75 .90 .95
537 0          45      23.72 8 10 12 24 32 39 40
lowest : 0 4 6 7 8, highest: 48 49 52 60 64
```

---

### temp : Temperature (celcius) Day 3



```
      n missing unique Mean .05 .10 .25 .50 .75 .90 .95
537 0          61      37.52 35.50 35.80 36.40 37.80 38.50 39.09 39.50
lowest : 32.50 34.00 34.09 34.90 35.00
highest: 40.20 40.59 40.90 41.00 41.20
```

---

### pafi : PaO2/(.01\*FiO2) Day 3



```
      n missing unique Mean .05 .10 .25 .50 .75 .90 .95
500 37          357      227.2 86.99 105.08 137.88 202.56 290.00 390.49 433.31
lowest : 45.00 48.00 53.33 54.00 55.00
highest: 574.00 595.12 640.00 680.00 869.38
```

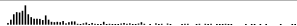
---

**alb : Serum Albumin Day 3**

|          | n     | missing | unique | Mean  | .05   | .10   | .25   | .50   | .75   | .90   | .95   |
|----------|-------|---------|--------|-------|-------|-------|-------|-------|-------|-------|-------|
|          | 346   | 191     | 34     | 2.668 | 1.700 | 1.900 | 2.225 | 2.600 | 3.100 | 3.400 | 3.800 |
| lowest : | 1.100 | 1.200   | 1.300  | 1.400 | 1.500 |       |       |       |       |       |       |
| highest: | 4.100 | 4.199   | 4.500  | 4.699 | 4.800 |       |       |       |       |       |       |

**bili : Bilirubin Day 3**

|          | n        | missing  | unique   | Mean     | .05      | .10    | .25    | .50    | .75    | .90    |
|----------|----------|----------|----------|----------|----------|--------|--------|--------|--------|--------|
|          | 386      | 151      | 88       | 2.678    | 0.3000   | 0.4000 | 0.6000 | 0.8999 | 2.0000 | 6.5996 |
|          |          |          | .95      | 13.1743  |          |        |        |        |        |        |
| lowest : | 0.09999  | 0.19998  | 0.29999  | 0.39996  | 0.50000  |        |        |        |        |        |
| highest: | 22.59766 | 30.00000 | 31.50000 | 35.00000 | 39.29688 |        |        |        |        |        |

**crea : Serum creatinine Day 3**

|          | n    | missing | unique | Mean  | .05    | .10    | .25    | .50    | .75    | .90    | .95    |
|----------|------|---------|--------|-------|--------|--------|--------|--------|--------|--------|--------|
|          | 537  | 0       | 84     | 2.232 | 0.6000 | 0.7000 | 0.8999 | 1.3999 | 2.5996 | 5.2395 | 7.3197 |
| lowest : | 0.3  | 0.4     | 0.5    | 0.6   | 0.7    |        |        |        |        |        |        |
| highest: | 10.4 | 10.6    | 11.2   | 11.6  | 11.8   |        |        |        |        |        |        |

### sod : Serum sodium Day 3



```
      n missing unique  Mean  .05  .10  .25  .50  .75  .90  .95
537  0          38      138.1 129 131 134 137 142 147 150
lowest : 118 120 121 126 127, highest: 156 157 158 168 175
```

---

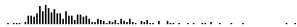
### ph : Serum pH (arterial) Day 3



```
      n missing unique  Mean  .05  .10  .25  .50  .75  .90  .95
500  37          49      7.416 7.270 7.319 7.380 7.420 7.470 7.510 7.529
lowest : 6.960 6.989 7.069 7.119 7.130
highest: 7.560 7.569 7.590 7.600 7.659
```

---

### glucose : Glucose Day 3



```
      n missing unique  Mean  .05  .10  .25  .50  .75  .90  .95
297  240          179      167.7 76.0 89.0 106.0 141.0 200.0 292.4 347.2
lowest : 30 42 52 55 68, highest: 446 468 492 576 598
```

---

**bun : BUN Day 3**

|          | n   | missing | unique | Mean  | .05                             | .10   | .25   | .50   | .75   | .90   | .95    |
|----------|-----|---------|--------|-------|---------------------------------|-------|-------|-------|-------|-------|--------|
|          | 304 | 233     | 100    | 38.91 | 8.00                            | 11.00 | 16.75 | 30.00 | 56.00 | 79.70 | 100.70 |
| lowest : | 1   | 3       | 4      | 5     | 6, highest: 123 124 125 128 146 |       |       |       |       |       |        |

**urine : Urine Output Day 3**

|          | n   | missing | unique | Mean | .05                                   | .10   | .25    | .50    | .75    | .90    | .95    |
|----------|-----|---------|--------|------|---------------------------------------|-------|--------|--------|--------|--------|--------|
|          | 303 | 234     | 262    | 2095 | 20.3                                  | 364.0 | 1156.5 | 1870.0 | 2795.0 | 4008.6 | 4817.5 |
| lowest : | 0   | 5       | 8      | 15   | 20, highest: 6865 6920 7360 7560 7750 |       |        |        |        |        |        |

**adlp : ADL Patient Day 3**

|           | n   | missing | unique | Mean  |   |   |   |   |   |
|-----------|-----|---------|--------|-------|---|---|---|---|---|
|           | 104 | 433     | 8      | 1.577 |   |   |   |   |   |
| Frequency |     | 0       | 1      | 2     | 3 | 4 | 5 | 6 | 7 |
|           | 51  | 19      | 7      | 6     | 4 | 7 | 8 | 2 |   |
| %         | 49  | 18      | 7      | 6     | 4 | 7 | 8 | 2 |   |

### adls : ADL Surrogate Day 3

|           | n   | missing | unique | Mean |    |    |    |    |
|-----------|-----|---------|--------|------|----|----|----|----|
|           | 392 | 145     | 8      | 1.86 |    |    |    |    |
| Frequency | 0   | 1       | 2      | 3    | 4  | 5  | 6  | 7  |
|           | 185 | 68      | 22     | 18   | 17 | 20 | 39 | 23 |
| %         | 47  | 17      | 6      | 5    | 4  | 5  | 10 | 6  |

---

### sfdm2 : Severe Functional Disability Month 2

|  | n   | missing | unique |
|--|-----|---------|--------|
|  | 468 | 69      | 5      |

no(M2 and SIP pres) (134, 29%), adl>=4 (>=5 if sur) (78, 17%)  
SIP>=30 (30, 6%), Coma or Intub (5, 1%), <2 mo. follow-up (221, 47%)

---

### adlsc : Imputed ADL Calibrated to Surrogate

|  | n   | missing | unique | Mean  | .05   | .10   | .25   | .50   | .75   | .90   | .95   |
|--|-----|---------|--------|-------|-------|-------|-------|-------|-------|-------|-------|
|  | 537 | 0       | 144    | 2.119 | 0.000 | 0.000 | 0.000 | 1.839 | 3.375 | 6.000 | 6.000 |

lowest : 0.0000 0.4948 0.4948 1.0000 1.1667  
highest: 5.7832 6.0000 6.3398 6.4658 7.0000

---

```
> # Show patterns of missing data
> plot(naclus(support[acute,])) # Figure 12
```

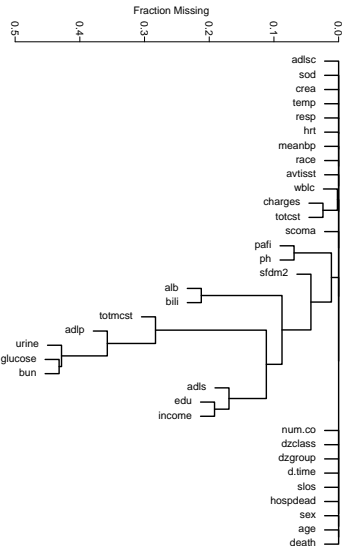
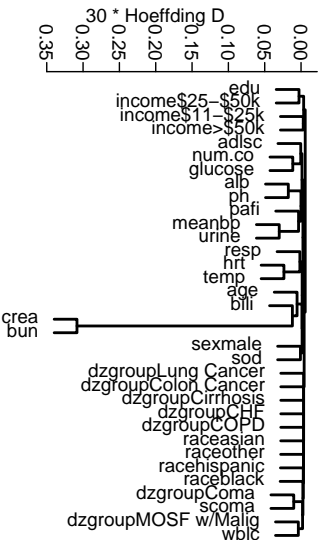


Figure 12: Cluster analysis showing which predictors tend to be missing on the same patients

Show associations between predictors using a general non-monotonic

measure of dependence (Hoeffding  $D$ ).

```
> ac <- support[acute,]
> ac$dzgroup <- ac$dzgroup[drop=TRUE]      # Remove unused levels
> attach(ac)
> vc <- varclus(~ age+sex+dzgroup+num.co+edu+income+scoma+race+
+               meanbp+wblc+hrt+resp+temp+pafi+alb+bili+crea+sod+
+               ph+glucose+bun+urine+adlsc, sim='hoeffding')
> plot(vc)                                  # Figure 13
```



**Figure 13:** Hierarchical clustering of potential predictors using Hoeffding  $D$  as a similarity measure. Categorical predictors are automatically expanded into dummy variables.

# Checking Adequacy of Log-Normal Accelerated Failure Time Model

```
> dd ← datadist(ac)
> # describe distributions of variables to Design
> options(datadist='dd')

> # Generate right-censored survival time variable
> years ← d.time/365.25
> units(years) ← 'Year'
> S ← Surv(years, death)

> # Show normal inverse Kaplan-Meier estimates
> # stratified by dzgroup
> survplot(survfit(S ~ dzgroup), conf='none',
+          fun=qnorm,logt=TRUE) # Figure 14
```

More stringent assessment of log-normal assumptions: check distribution of residuals from an adjusted model:

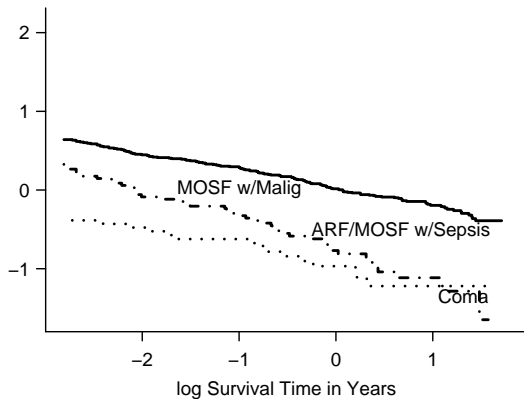


Figure 14:  $\Phi^{-1}(S_{KM}(t))$  stratified by `dzgroup`. Linearity and semi-parallelism indicate a reasonable fit to the log-normal accelerated failure time model with respect to one predictor.

```

> f ← psm(S ~ dzgroup + rcs(age,5) + rcs(meanbp,5),
          dist='lognormal', y=TRUE) # dist='gaussian' for S+ 2000
> r ← resid(f)

> par(mfrow=c(2,2))
> survplot(r, dzgroup, label.curve=FALSE)
> survplot(r, age, label.curve=FALSE)
> survplot(r, meanbp, label.curve=FALSE)
> random.number ← runif(length(age))
> survplot(r, random.number, label.curve=FALSE) # Figure 15

```

The fit for `dzgroup` is not great but overall fit is good.

Remove from consideration predictors that are missing in  $> 0.2$  of the patients. Many of these were only collected for the second phase of SUPPORT.

Of those continuous variables to be included in the model, find which ones have enough potential predictive power to justify allowing for non-linear relationships, which spend more d.f. For each variable compute

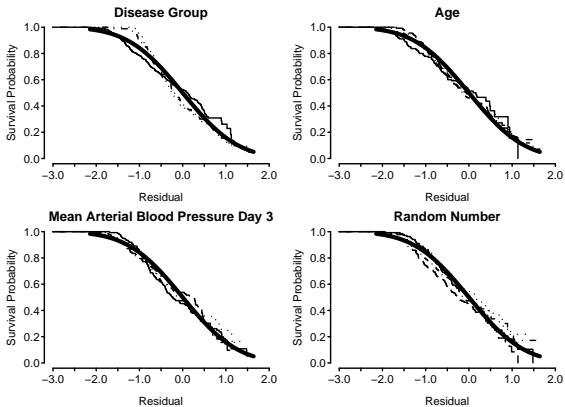


Figure 15: Kaplan-Meier estimates of distributions of normalized, right-censored residuals from the fitted log-normal survival model. Residuals are stratified by important variables in the model (by quartiles of continuous variables), plus a random variable to depict the natural variability. Theoretical standard Gaussian distributions of residuals is shown with a thick solid line.

Spearman  $\rho^2$  based on multiple linear regression of  $\text{rank}(x)$ ,  $\text{rank}(x)^2$  and the survival time, truncating survival time at the shortest follow-up for survivors (356 days).

```
> shortest.follow.up ← min(d.time[death==0], na.rm=TRUE)
> d.timet ← pmin(d.time, shortest.follow.up)
```

```
> w ← spearman2(d.timet ~ age + num.co + scoma + meanbp +
+             hrt + resp + temp + crea + sod + adlsc +
+             wblc + pafi + ph + dzgroup + race, p=2)
> plot(w)           # Figure 16
```

```
> # Compute number of missing values per variable
> sapply(llist(age,num.co,scoma,meanbp,hrt,resp,temp,crea,sod,adlsc
+             wblc,pafi,ph), function(x) sum(is.na(x)))
> # Can also do nplot(naclus(support[acute,]))
```

```
age num.co scoma meanbp hrt resp temp crea sod adlsc wblc pafi ph
  0      0      0      0   0   0   0   0   0   0      5  37 37
```

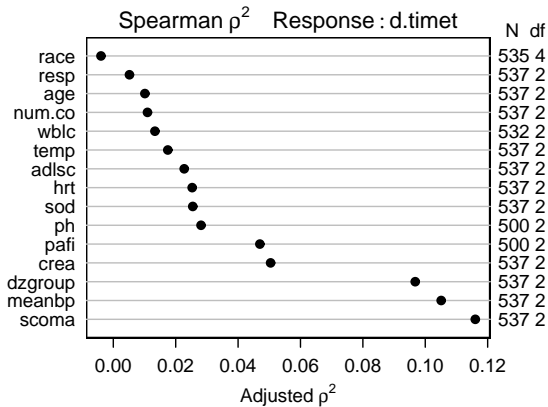


Figure 16: Generalized Spearman  $\rho^2$  rank correlation between predictors and truncated survival time

```

> # Can also use the Hmisc naclus and naplot functions to do this

> # Impute missing values with normal or modal values
> wblc.i ← impute(wblc.i, 9)
> pafi.i ← impute(pafi.i, 333.3)
> ph.i ← impute(ph.i, 7.4)
> race2 ← race
> levels(race2) ← list(white='white',other=levels(race)[-1])
> race2[is.na(race2)] ← 'white'
> dd ← datadist(dd, wblc.i, pafi.i, ph.i, race2)

```

Do a formal redundancy analysis using more than pairwise associations, and allow for non-monotonic transformations in predicting each predictor from all other predictors. This analysis requires missing values to be imputed so as to not greatly reduce the sample size.

```

> redun(~ crea + age + sex + dzgroup + num.co + scoma + adlsc + race
+       meanbp + hrt + resp + temp + sod + wblc.i + pafi.i + ph.i, nk

```

## Redundancy Analysis

```
redun(formula = ~crea + age + sex + dzgroup + num.co + scoma +  
      adlsc + race2 + meanbp + hrt + resp + temp + sod + wblc.i +  
      pafi.i + ph.i, nk = 4)
```

n: 537 p: 16 nk: 4

Number of NAs: 0

Transformation of target variables forced to be linear

R-squared cutoff: 0.9 Type: ordinary

R<sup>2</sup> with which each variable can be predicted from all other variables:

|       |       |       |         |        |       |       |       |        |
|-------|-------|-------|---------|--------|-------|-------|-------|--------|
| crea  | age   | sex   | dzgroup | num.co | scoma | adlsc | race2 | meanbp |
| 0.133 | 0.246 | 0.132 | 0.451   | 0.147  | 0.418 | 0.153 | 0.151 | 0.178  |
| resp  | temp  | sod   | wblc.i  | pafi.i | ph.i  |       |       |        |
| 0.131 | 0.197 | 0.135 | 0.093   | 0.143  | 0.171 |       |       |        |

No redundant variables

## Better approach to gauging predictive potential and allocating d.f.:

- Allow all continuous variables to have a the maximum number of knots entertained, in a log-normal survival model
- Must use imputation to avoid losing data
- Fit a “saturated” main effects model
- Makes full use of censored data
- Had to limit to 4 knots, force `scoma` to be linear, and omit `ph.i` to avoid singularity

```
> k ← 4
> f ← psm(S ~ rcs(age,k)+sex+dzgroup+pol(num.co,2)+scoma+
  pol(adlsc,2)+race+rcs(meanbp,k)+rcs(hrt,k)+rcs(resp,k)+
  rcs(temp,k)+rcs(crea,3)+rcs(sod,k)+rcs(wblc.i,k)+
  rcs(pafi.i,k), dist='lognormal')
> plot(anova(f)) # Figure 17
```

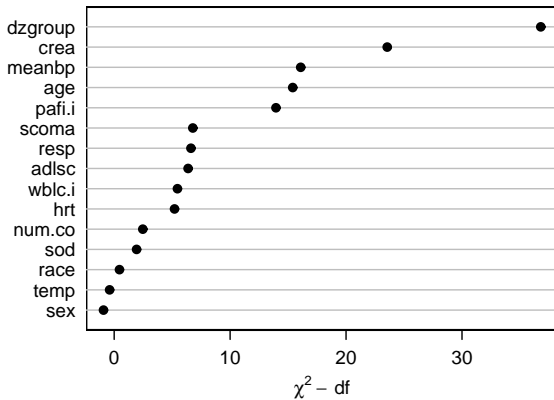


Figure 17: Partial  $\chi^2$  statistics for association of each predictor with response from saturated main effects model, penalized for d.f.

- Figure 17 properly blinds the analyst to the form of effects (tests of linearity).
- Fit a log-normal survival model with number of parameters corresponding to nonlinear effects determined from Figure 17. For the most promising predictors, five knots can be allocated, as there are fewer singularity problems once less promising predictors are simplified.

```
> f ← psm(S ~ rcs(age,5)+sex+dzgroup+num.co+
+         scoma+pol(adlsc,2)+race2+rcs(meanbp,5)+
+         rcs(hrt,3)+rcs(resp,3)+temp+
+         rcs(crea,4)+sod+rcs(wblc.i,3)+rcs(pafi.i,4),
+         dist='lognormal') # 'gaussian' for S+ 2000
> f
```

Parametric Survival Model: Log Normal Distribution

| Obs | Events | Model L.R. | d.f. | P | R2   |
|-----|--------|------------|------|---|------|
| 537 | 356    | 236.83     | 30   | 0 | 0.59 |
| .   | .      | .          | .    | . | .    |

# Summarizing the Fitted Model

- Plot the shape of the effect of each predictor on log survival time.
- All effects centered: can be placed on common scale
- Wald  $\chi^2$  statistics, penalized for d.f., plotted in descending order

```
> par(mfrow=c(3,5))  
> plot(f, ref.zero=TRUE, ylim=c(-3,2), lwd.conf=.35) # Figure 18
```

```
> plot(anova(f)) # Figure 19  
> anova(f)
```

```
> options(digits=3)  
> plot(summary(f), log=TRUE) # Figure 20
```

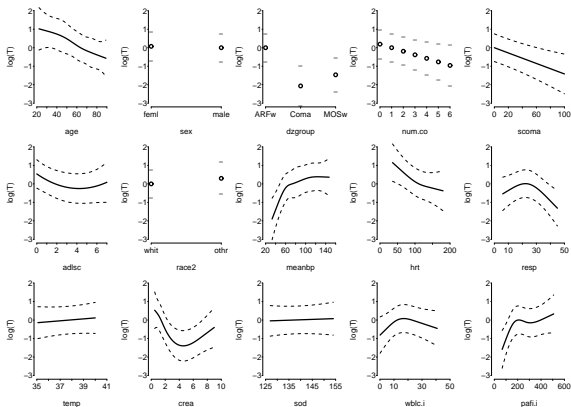


Figure 18: Effect of each predictor on log survival time. Predicted values have been centered so that predictions at predictor reference values are zero. Pointwise 0.95 confidence bands are also shown. As all Y-axes have the same scale, it is easy to see which predictors are strongest.

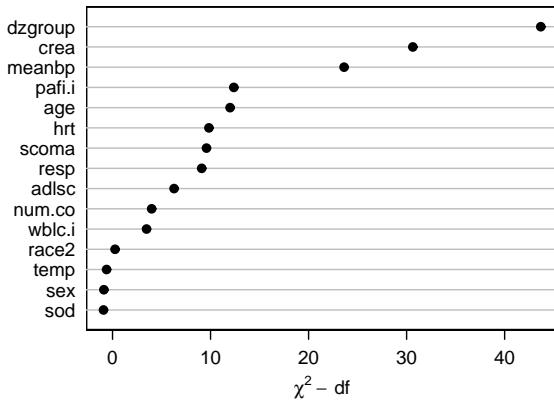


Figure 19: Contribution of variables in predicting survival time in log-normal model

Table 1: Wald Statistics for S

|                        | $\chi^2$ | <i>d.f.</i> | <i>P</i> |
|------------------------|----------|-------------|----------|
| age                    | 15.99    | 4           | 0.0030   |
| <i>Nonlinear</i>       | 0.23     | 3           | 0.9722   |
| sex                    | 0.11     | 1           | 0.7354   |
| dzgroup                | 45.69    | 2           | < 0.0001 |
| num.co                 | 4.99     | 1           | 0.0255   |
| scoma                  | 10.58    | 1           | 0.0011   |
| adlsc                  | 8.28     | 2           | 0.0159   |
| <i>Nonlinear</i>       | 3.31     | 1           | 0.0691   |
| race2                  | 1.26     | 1           | 0.2624   |
| meanbp                 | 27.62    | 4           | < 0.0001 |
| <i>Nonlinear</i>       | 10.51    | 3           | 0.0147   |
| hrt                    | 11.83    | 2           | 0.0027   |
| <i>Nonlinear</i>       | 1.04     | 1           | 0.3090   |
| resp                   | 11.10    | 2           | 0.0039   |
| <i>Nonlinear</i>       | 8.56     | 1           | 0.0034   |
| temp                   | 0.39     | 1           | 0.5308   |
| crea                   | 33.63    | 3           | < 0.0001 |
| <i>Nonlinear</i>       | 21.27    | 2           | < 0.0001 |
| sod                    | 0.08     | 1           | 0.7792   |
| wbhc.i                 | 5.47     | 2           | 0.0649   |
| <i>Nonlinear</i>       | 5.46     | 1           | 0.0195   |
| pafi.i                 | 15.37    | 3           | 0.0015   |
| <i>Nonlinear</i>       | 6.97     | 2           | 0.0307   |
| <b>TOTAL NONLINEAR</b> | 60.48    | 14          | < 0.0001 |
| <b>TOTAL</b>           | 261.47   | 30          | < 0.0001 |

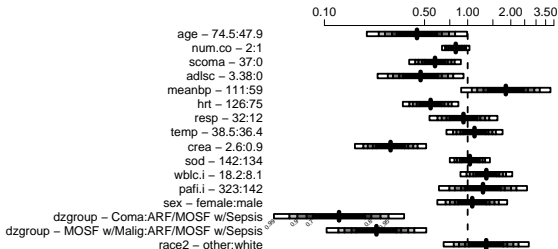


Figure 20: Estimated survival time ratios for default settings of predictors. For example, when age changes from its lower quartile to the upper quartile (47.9y to 74.5y), median survival time decreases by more than half. Different shaded areas of bars indicate different confidence levels, ranging from 0.7 to 0.99.

# Internal Validation of the Fitted Model Using the Bootstrap

Validate indexes describing the fitted model.

```
> # First add data to model fit so bootstrap can re-sample
> # from the data
> g ← update(f, x=TRUE, y=TRUE)
> set.seed(717)
> validate(g, B=120, dxy=TRUE)
```

|           | index.orig | training | test   | optimism | index.corrected | n   |
|-----------|------------|----------|--------|----------|-----------------|-----|
| Dxy       | 0.4852     | 0.5126   | 0.459  | 0.054    | 0.432           | 120 |
| R2        | 0.5940     | 0.6584   | 0.537  | 0.122    | 0.472           | 120 |
| Intercept | 0.0000     | 0.0000   | -0.056 | 0.056    | -0.056          | 120 |
| Slope     | 1.0000     | 1.0000   | 0.900  | 0.100    | 0.900           | 120 |
| D         | 0.4788     | 0.5483   | 0.422  | 0.126    | 0.353           | 120 |
| U         | -0.0041    | -0.0041  | 0.106  | -0.110   | 0.106           | 120 |
| Q         | 0.4829     | 0.5524   | 0.316  | 0.236    | 0.247           | 120 |

- From  $D_{xy}$  and  $R^2$  there is a moderate amount of overfitting.
- Slope shrinkage factor (0.90) is not troublesome
- Almost unbiased estimate of future predictive discrimination on similar patients is the corrected  $D_{xy}$  of 0.43.

Validate predicted 1-year survival probabilities. Use Kaplan-Meier estimates obtained by stratifying patients by the predicted probability, with at least 60 patients per group.

```
> set.seed(717)
> cal ← calibrate(g, u=1, m=60, B=120)
> plot(cal) # Figure 21
```

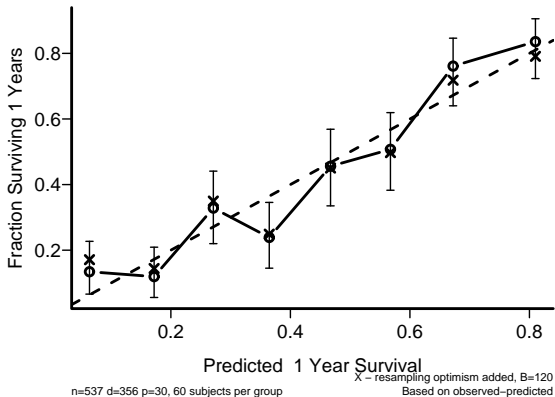


Figure 21: Bootstrap validation of calibration curve. Dots represent apparent calibration accuracy;  $\times$  are bootstrap estimates corrected for overfitting.

## Approximating the Full Model

The fitted log-normal model is perhaps too complex for routine use and for routine data collection. Let us develop a simplified model that can predict the predicted values of the full model with high accuracy ( $R^2 = 0.96$ ). The simplification is done using a fast backward stepdown against the full model predicted values.

```
> Z ← predict(f)      # X*beta hat
> a ← ols(Z ~ rcs(age,5)+sex+dzgroup+num.co+
+         scoma+pol(adlsc,2)+race2+
+         rcs(meanbp,5)+rcs(hrt,3)+rcs(resp,3)+
+         temp+rcs(crea,4)+sod+rcs(wblc.i,3)+
+         rcs(pafi.i,4), sigma=1)
> # sigma=1 is used to prevent sigma hat from being zero when
> # R2=1.0 since we start out by approximating Z with all
> # component variables
> fastbw(a, aics=10000)  # fast backward stepdown
```

| Deleted | Chi-Sq | d.f. | P     | Residual | d.f. | P    | AIC    | R2   |
|---------|--------|------|-------|----------|------|------|--------|------|
| sod     | 0.43   | 1    | 0.512 | 0.43     | 1    | 0.51 | -1.6   | 1.00 |
| sex     | 0.57   | 1    | 0.451 | 1.00     | 2    | 0.61 | -3.0   | 1.00 |
| temp    | 2.20   | 1    | 0.138 | 3.20     | 3    | 0.36 | -2.8   | 1.00 |
| race2   | 6.81   | 1    | 0.009 | 10.01    | 4    | 0.04 | 2.0    | 0.99 |
| wblc.i  | 29.52  | 2    | 0.000 | 39.53    | 6    | 0.00 | 27.5   | 0.98 |
| num.co  | 30.84  | 1    | 0.000 | 70.36    | 7    | 0.00 | 56.4   | 0.96 |
| resp    | 54.18  | 2    | 0.000 | 124.55   | 9    | 0.00 | 106.5  | 0.92 |
| adlsc   | 52.46  | 2    | 0.000 | 177.00   | 11   | 0.00 | 155.0  | 0.89 |
| pafi.i  | 66.78  | 3    | 0.000 | 243.79   | 14   | 0.00 | 215.8  | 0.85 |
| scoma   | 78.07  | 1    | 0.000 | 321.86   | 15   | 0.00 | 291.9  | 0.80 |
| hrt     | 83.17  | 2    | 0.000 | 405.02   | 17   | 0.00 | 371.0  | 0.75 |
| age     | 68.08  | 4    | 0.000 | 473.10   | 21   | 0.00 | 431.1  | 0.71 |
| crea    | 314.47 | 3    | 0.000 | 787.57   | 24   | 0.00 | 739.6  | 0.52 |
| meanbp  | 403.04 | 4    | 0.000 | 1190.61  | 28   | 0.00 | 1134.6 | 0.27 |
| dzgroup | 441.28 | 2    | 0.000 | 1631.89  | 30   | 0.00 | 1571.9 | 0.00 |

```

> f.approx ← ols(Z ~ dzgroup + rcs(meanbp,5) + rcs(crea,4) + rcs(a
      rcs(hrt,3) + scoma + rcs(pafi.i,4) + pol(adlsc,2)+
      rcs(resp,3), x=TRUE)
> f.approx$stats

```

| n      | Model L.R. | d.f.  | R2   | Sigma |
|--------|------------|-------|------|-------|
| 537.00 | 1688.22    | 23.00 | 0.96 | 0.37  |

- Estimate variance–covariance matrix of the coefficients of reduced model
- This covariance matrix does not include the scale parameter

```

> V ← Varcov(f,regcoef.only=TRUE)           # var(full model)
> X ← g$x                                   # full model design
> x ← f.approx$x                           # approx. model design
> w ← solve(t(x) %*% x, t(x)) %*% X        # contrast matrix
> v ← w %*% V %*% t(w)

```

Compare variance estimates (diagonals of  $v$ ) with variance estimates from a reduced model that is fitted against the actual outcomes.

```

> f.sub ← psm(S ~ dzgroup + rcs(meanbp,5) + rcs(crea,4) + rcs(age,

```

```
+          rcs(hrt,3) + scoma + rcs(pafi.i,4) + pol(adlsc,2)+  
+          rcs(resp,3), dist='lognormal') # 'gaussian' for S+  
  
> diag(v)/diag(Varcov(f.sub,regcoef.only=TRUE))
```

The ratios ranged from 0.978 to 0.982.

```
> f.approx$var ← v  
> anova(f.approx, test='Chisq', ss=FALSE)
```

Equation for simplified model:

```
> # Typeset mathematical form of approximate model  
> w ← latex(f.approx)
```

Table 2: Wald Statistics for Z

|                        | $\chi^2$ | <i>d.f.</i> | <i>P</i> |
|------------------------|----------|-------------|----------|
| dzgroup                | 2218.73  | 2           | < 0.0001 |
| meanbp                 | 1167.98  | 4           | < 0.0001 |
| <i>Nonlinear</i>       | 385.01   | 3           | < 0.0001 |
| crea                   | 1591.35  | 3           | < 0.0001 |
| <i>Nonlinear</i>       | 959.71   | 2           | < 0.0001 |
| age                    | 778.00   | 4           | < 0.0001 |
| <i>Nonlinear</i>       | 14.18    | 3           | 0.0027   |
| hrt                    | 413.17   | 2           | < 0.0001 |
| <i>Nonlinear</i>       | 16.55    | 1           | < 0.0001 |
| scoma                  | 383.38   | 1           | < 0.0001 |
| pafi.i                 | 556.37   | 3           | < 0.0001 |
| <i>Nonlinear</i>       | 268.81   | 2           | < 0.0001 |
| adlsc                  | 402.21   | 2           | < 0.0001 |
| <i>Nonlinear</i>       | 116.31   | 1           | < 0.0001 |
| resp                   | 395.03   | 2           | < 0.0001 |
| <i>Nonlinear</i>       | 295.65   | 1           | < 0.0001 |
| <b>TOTAL NONLINEAR</b> | 2335.64  | 13          | < 0.0001 |
| <b>TOTAL</b>           | 11384.50 | 23          | < 0.0001 |

$$E(Z) = X\beta, \text{ where}$$

$$\begin{aligned}
X\hat{\beta} = & -2.509 \\
& -1.941\{\text{Coma}\} - 1.753\{\text{MOSF w/Malig}\} \\
& +0.06798\text{meanbp} - 3.077 \times 10^{-5}(\text{meanbp} - 41.8)_+^3 + 7.896 \times 10^{-5}(\text{meanbp} - 61)_+^3 \\
& -4.909 \times 10^{-5}(\text{meanbp} - 73)_+^3 + 2.607 \times 10^{-6}(\text{meanbp} - 108.6)_+^3 \\
& -1.701 \times 10^{-6}(\text{meanbp} - 135)_+^3 \\
& -0.5528\text{crea} - 0.2294(\text{crea} - 0.6)_+^3 + 0.4496(\text{crea} - 1.100)_+^3 \\
& -0.2332(\text{crea} - 1.94)_+^3 + 0.01306(\text{crea} - 7.32)_+^3 \\
& -0.01646\text{age} - 1.132 \times 10^{-5}(\text{age} - 28.49)_+^3 + 4.055 \times 10^{-5}(\text{age} - 49.52)_+^3 \\
& -2.146 \times 10^{-5}(\text{age} - 63.67)_+^3 - 2.677 \times 10^{-5}(\text{age} - 72.66)_+^3 + 1.901 \times 10^{-5}(\text{age} - 85.56)_+^3 \\
& -0.01361\text{hrt} + 6.093 \times 10^{-7}(\text{hrt} - 60)_+^3 - 1.681 \times 10^{-6}(\text{hrt} - 111)_+^3 \\
& +1.072 \times 10^{-6}(\text{hrt} - 140)_+^3 - 0.01354 \text{scoma} \\
& +0.01607\text{pafi.i} - 4.769 \times 10^{-7}(\text{pafi.i} - 87.99)_+^3 + 9.113 \times 10^{-7}(\text{pafi.i} - 166.7)_+^3 \\
& -5.021 \times 10^{-7}(\text{pafi.i} - 276.2)_+^3 + 6.764 \times 10^{-8}(\text{pafi.i} - 425.6)_+^3 \\
& -0.36925 \text{adlsc} + 0.04087 \text{adlsc}^2
\end{aligned}$$

$$+0.03941\text{resp} - 9.106 \times 10^{-5}(\text{resp} - 10)_+^3 + 0.0001760(\text{resp} - 24)_+^3 \\ - 8.499 \times 10^{-5}(\text{resp} - 39)_+^3$$

and  $\{c\} = 1$  if subject is in group  $c$ , 0 otherwise;  $(x)_+ = x$  if  $x > 0$ , 0 otherwise.

Nomogram for predicting median and mean survival time, based on approximate model:

```
> # Derive S functions that express mean and quantiles
> # of survival time for specific linear predictors
> # analytically
> expected.surv ← Mean(f)
> quantile.surv ← Quantile(f)
> expected.surv

function(lp, parms = structure(.Data = 0.802352037606488,
                              .Names = "Log(scale)"),
        transform = structure(.Data = "log", .Names = "link"))
{
  names(parms) ← NULL
```

```

switch(transform,
      identity = lp,
      log = exp(lp + exp(2 * parms)/2),
      stop(paste(transform, "not implemented")))
}

> quantile.surv

function(q = 0.5, lp,
      parms = structure(.Data = 0.802352037606488,
      .Names = "Log(scale)"),
      transform = structure(.Data = "log",
      .Names = "link"))
{
names(parms) ← NULL
inv ← glm.links["inverse", transform]$inverse
f ← function(lp, q, parms)
lp + exp(parms) * qnorm(q)
names(q) ← format(q)
drop(inv(outer(lp, q, FUN = f, parms = parms)))
}

```

```

> median.surv ← function(x) quantile.surv(lp=x)

> # Improve variable labels for the nomogram
> f.approx ← Newlabels(f.approx, c('Disease Group', 'Mean Arterial
+       'Creatinine', 'Age', 'Heart Rate', 'SUPPORT Coma Score',
+       'PaO2/(.01*FiO2)', 'ADL', 'Resp. Rate'))
> nomogram(f.approx,
+       pafi.i=c(0, 50, 100, 200, 300, 500, 600, 700, 800, 900),
+       fun=list('Median Survival Time'=median.surv,
+               'Mean Survival Time' =expected.surv),
+       fun.at=c(.1, .25, .5, 1, 2, 5, 10, 20, 40),
+       cex.var=1, cex.axis=.75, lmgp=.25)
> # Figure 22

```

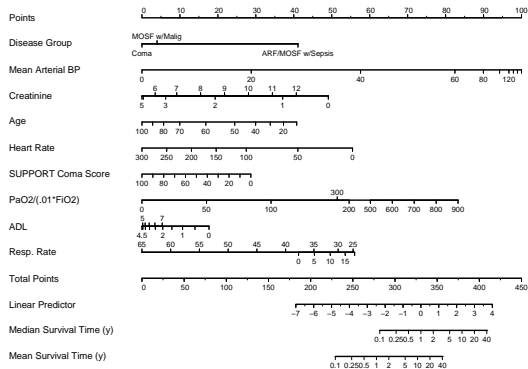


Figure 22: Nomogram for predicting median and mean survival time, based on approximation of full model

## S Packages and Functions Used

| Packages | Purpose   | Functions   |
|----------|---|---|
| Hmisc    | Miscellaneous functions                                 | describe,ecdf,naclus,<br>varclus,l1ist,spearman2<br>describe,impute,latex                                     |
| Design   | Modeling, validation,<br>Model presentation<br>graphics | datadist,survplot,Newlabels,<br>psm,validate,calibrate,<br>Function,rcs,ols,fastbw,<br>Mean,Quantile,nomogram |

Note: All packages are available from CRAN

\*

## Bibliography

- [1] K. Akazawa, T. Nakamura, and Y. Palesch. Power of logrank test and Cox regression model in clinical trials with heterogeneous samples. *Stat Med*, 16:583–597, 1997. []
- [2] D. G. Altman and C. J. Doré. Randomization and baseline comparison in clinical trials. *Lancet*, 335:149–153, 1990. []
- [3] P. K. Andersen, J. P. Klein, and M. Zhang. Testing for centre effects in multi-centre survival studies: A monte carlo comparison of fixed and random effects tests. *Stat Med*, 18:1489–1500, 1999. []
- [4] C. B. Begg. Significance tests of covariate imbalance in clinical trials. *Controlled Clin Trials*, 11:223–225, 1990. []
- [5] Committee for Proprietary Medicinal Products. Points to consider on adjustment for baseline covariates. *Stat Med*, 23:701–709, 2004. [10]
- [6] C. Cox, H. Chu, M. F. Schneider, and A. Muñoz. Parametric survival analysis and taxonomy of hazard functions for the generalized gamma distribution. *Stat Med*, 26:4352–4374, 2007. [31]
- [7] CPMP Working Party. Biostatistical methodology in clinical trials in applications for marketing authorizations for medicinal products. *Stat Med*, 14:1659–1682, 1995. []
- [8] W. D. Dupont. *Statistical Modeling for Biomedical Researchers*. Cambridge University Press, Cambridge, UK, second edition, 2008. [106]
- [9] D. Edwards. On model pre-specification in confirmatory randomized studies. *Stat Med*, 18:771–785, 1999. [10, 14]

- [10] I. Ford and J. Norrie. The role of covariates in estimating treatment effects and risk in long-term clinical trials. *Stat Med*, 21:2899–2908, 2002. []
- [11] I. Ford, J. Norrie, and S. Ahmadi. Model inconsistency, illustrated by the Cox proportional hazards model. *Stat Med*, 14:735–746, 1995. []
- [12] M. Gail, S. Wieand, and S. Piantadosi. Biased estimates of treatment effect in randomized experiments with nonlinear regressions and omitted covariates. *Biometrika*, 71:431–444, 1984. []
- [13] M. H. Gail. Adjusting for covariates that have the same distribution in exposed and unexposed cohorts. In S. H. Moolgavkar and R. L. Prentice, editors, *Modern Statistical Methods in Chronic Disease Epidemiology*, pages 3–18. Wiley, New York, 1986. []
- [14] M. H. Gail, W. Y. Tan, and S. Piantadosi. Tests for no treatment effect in randomized clinical trials. *Biometrika*, 75:57–64, 1988. []
- [15] F. E. Harrell. *Regression Modeling Strategies, with Applications to Linear Models, Survival Analysis and Logistic Regression*. Springer, New York, 2001. []
- [16] 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. *Stat Med*, 15:361–387, 1996. []
- [17] W. W. Hauck, S. Anderson, and S. M. Marcus. Should we adjust for covariates in nonlinear regression analyses of randomized trials? *Controlled Clin Trials*, 19:249–256, 1998. [8, 9]
- [18] W. W. Hauck, J. M. Neuhaus, J. D. Kalbfleisch, and S. Anderson. A consequence of omitted covariates when estimating odds ratios. *J Clin Epi*, 44:77–81, 1991. []

- [19] A. V. Hernández, E. W. Steyerberg, and J. D. F. Habbema. Covariate adjustment in randomized controlled trials with dichotomous outcomes increases statistical power and reduces sample size requirements. *J Clin Epi*, 57:454–460, 2004. []
- [20] J. E. Herndon and F. E. Harrell. The restricted cubic spline hazard model. *Comm Stat Th Meth*, 19:639–663, 1990. [31]
- [21] J. E. Herndon and F. E. Harrell. The restricted cubic spline as baseline hazard in the proportional hazards model with step function time-dependent covariables. *Stat Med*, 14:2119–2129, 1995. [31]
- [22] International Conference on Harmonisation. Guideline E-9: Statistical Principles for Clinical Trials, 1998. []
- [23] A. Källén. Treatment-by-center interaction: What is the issue? *Drug Info J*, 31:927–936, 1997. []
- [24] W. A. Knaus, F. E. Harrell, C. J. Fisher, D. P. Wagner, S. M. Opan, J. C. Sadoff, E. A. Draper, C. A. Walawander, K. Conboy, and T. H. Grasela. The clinical evaluation of new drugs for sepsis: A prospective study design based on survival analysis. *JAMA*, 270:1233–1241, 1993. []
- [25] W. A. Knaus, F. E. Harrell, J. Lynn, L. Goldman, R. S. Phillips, A. F. Connors, N. V. Dawson, W. J. Fulkerson, R. M. Califf, N. Desbiens, P. Layde, R. K. Oye, P. E. Bellamy, R. B. Hakim, and D. P. Wagner. The SUPPORT prognostic model: Objective estimates of survival for seriously ill hospitalized adults. *Ann Int Med*, 122:191–203, 1995. [54]
- [26] C. Kooperberg, C. J. Stone, and Y. K. Truong. Hazard regression. *J Am Stat Assoc*, 90:78–94, 1995. [31]

- [27] S. W. Lagakos and D. A. Schoenfeld. Properties of proportional hazards score tests under misspecified regression models. *Biometrics*, 40:1037–1048, 1984. []
- [28] J. A. Lewis. Statistical issues in the regulation of medicines. *Stat Med*, 14:127–136, 1995. []
- [29] D. B. Macnaughton. Which sums of squares are best in unbalanced analysis of variance? <http://www.matstat.com/ss/>, 1998. []
- [30] T. M. Morgan. Omitting covariates from the proportional hazards model. *Biometrics*, 42:993–995, 1986. []
- [31] J. M. Neuhaus. Estimation efficiency with omitted covariates in generalized linear models. *J Am Stat Assoc*, 93:1124–1129, 1998. [9]
- [32] M. C. Pike. A method of analysis of certain class of experiments in carcinogenesis. *Biometrics*, 22:142–161, 1966. [47]
- [33] S. J. Pocock, S. E. Assmann, L. E. Enos, and L. E. Kasten. Subgroup analysis, covariate adjustment and baseline comparisons in clinical trial reporting: current practice and problems. *Stat Med*, 21:2917–2930, 2002. []
- [34] G. M. Raab, S. Day, and J. Sales. How to select covariates to include in the analysis of a clinical trial. *Controlled Clin Trials*, 21:330–342, 2004. [9]
- [35] L. D. Robinson and N. P. Jewell. Some surprising results about covariate adjustment in logistic regression models. *International Statistical Review*, 59:227–240, 1991. []
- [36] D. J. Sargent and J. S. Hodges. A hierarchical model method for subgroup analysis of time-to-event data in the Cox regression setting. Presented at the Joint Statistical Meetings, Chicago, 1996. []

- [37] D. A. Schoenfeld. Sample size formulae for the proportional hazards regression model. *Biometrics*, 39:499–503, 1983. []
- [38] S. Senn. Testing for baseline balance in clinical trials. *Stat Med*, 13:1715–1726, 1994. []
- [39] S. Senn. *Statistical Issues in Drug Development*. Wiley, Chichester, England, 1997. []
- [40] S. Senn. Controversies concerning randomization and additivity in clinical trials. *Stat Med*, 23:3729–3753, 2004. [3, 15, 16]
- [41] A. L. Siqueira and J. M. G. Taylor. Treatment effects in a logistic model involving the Box-Cox transformation. *J Am Stat Assoc*, 94:240–246, 1999. [14]
- [42] E. W. Steyerberg. *Clinical Prediction Models*. Springer, New York, 2009. [, 106]
- [43] E. W. Steyerberg, P. M. M. Bossuyt, and K. L. Lee. Clinical trials in acute myocardial infarction: Should we adjust for baseline characteristics? *Am Heart J*, 139:745–751, 2000. Editorial, pp. 761-763. []
- [44] J. W. Tukey. Tightening the clinical trial. *Controlled Clin Trials*, 14:266–285, 1993. [9]
- [45] P. Verweij and H. C. van Houwelingen. Penalized likelihood in Cox regression. *Stat Med*, 13:2427–2436, 1994. []

R and S-PLUS functions written by FE Harrell are freely available from `biostat.mc.vanderbilt.edu/rms`. R packages are on CRAN.

To obtain a 588-page book with detailed examples and case studies and notes on the theory and applications of survival analysis, logistic regression, and linear models, order REGRESSION MODELING STRATEGIES *with Applications to Linear Models, Logistic Regression, and Survival Analysis* by FE Harrell from Springer NY (2001). Steyerberg<sup>42</sup> and Dupont<sup>8</sup> are excellent texts for accompanying the book.

To obtain a glossary of statistical terms and other handouts related to diagnostic and prognostic modeling, point your Web browser to [biostat.mc.vanderbilt.edu/ClinStat](http://biostat.mc.vanderbilt.edu/ClinStat).