Problems and Potentials in Modeling Survival

L. Richard Smith, Ph.D.
Frank E. Harrell, Jr., Ph.D.
Lawrence H. Muhlbaier, Ph.D.

Introduction

The three general goals of survival analysis are to: compare survival curves among two or more therapies, elucidate the factors that influence survival and estimate their effects, and estimate the future survival for an individual or for a population. In medical effectiveness research, these three components are inextricably linked. Naftel (in this volume) presents a compelling framework for accomplishing these goals through the use of a powerful parametric survival model. Although his discussion centers on a particular parametric model, the principles propounded are relevant to all forms of survival modeling.

This paper will discuss treatment comparison, covariate selection, and prediction in the context of any survival model, in addition to the strengths and weaknesses of the multiphase model and some implications of survival analysis. A brief appendix describes the multiphase model.

Comparing Treatment Effects

Comparing two treatments is relatively easy if both treatments are from the same underlying hazard function family. Then the underlying hazard function can be modified by covariates in a proportional hazards manner or through direct modification of one or more of the parameters of the hazard function (see appendix). If each treatment has a different form of hazard function, there is no generally accepted method. If a different hazard function is assumed for each treatment, one can only easily test for factors affecting treatment differences by stratifying on treatment and then testing for covariate-by-treatment interactions (Thall and Lachin, 1986; also see the appendix to Califf, Harrell, Lee, and others, 1989, for a description).

Naftel (in this volume) uses an alternative strategy. The model is fit to each treatment's data. If the underlying hazard functions are different, that is prima facie evidence that the treatments have different time-varying outcomes. The question then becomes, "At what times do the survival curves differ?" The predicted survival and 70-percent confidence limits are produced for each treatment for an individual patient based on the patient's particular covariate values. The two survival curves are then compared across any time points of interest. Where the curves are distinctly different—that is, where the confidence limits are widely separated or, equivalently, where the 95-percent confidence interval for the survival difference does not include zero—the treatments are declared different.

A classical statistician would find this approach troublesome. Examining each time point for significance may yield spurious conclusions due to the multiplicity of tests. The counterargument is that we have already found the treatments to be different; we are now only interested in where they are "practically" different. The multiplicity-of-tests problem is mitigated by several factors. First, the model tends to produce "smooth" estimates of the prediction curves, and thus differences between adjacent time points tend to be similar; there are no erratic adjacent differences. A further mitigating factor is the knowledge of the disease and its treatment and the understanding of modeling process the physician and statistician bring to the problem at hand. The knowledgeable investigator is unlikely to make too much of an isolated difference.

Covariates

An area not discussed at length in Naftel's paper (in this volume) but essential in any survival analysis is the difficult problem of choosing appropriate covariates. Four general areas relate to covariates: selection and re-

Dr. Smith is Assistant Professor, Dr. Harrell is Associate Professor, and Dr. Muhlbaier is Research Assistant Professor; all are with the Division of Biometry and Medical Informatics, Department of Community and Family Medicine, Duke University Medical Center.

liability, composite covariates or indices, the functional form of the covariate, and the number of covariates.

Selection and reliability. In most situations, many covariates influence outcome in addition to treatment. One common method of determining relevant covariates is through the use of a stepwise regression method. This method is systematic and objective, but as in all other aspects of the modeling process, there are dangers. The process of choosing the final set of covariates requires many tests of significance—the multiple comparison problem—that may lead to a large Type I error.

To understand the stability of the variable selection process, Efron and Gong (1983) proposed "bootstrap" methods, in which the data set is sampled with replacement m times, where m is some suitably large number. As an example, one of us (Harrell) ran a simulation for selection of covariates relating to time until death on a population of critically ill patients (N = 984, deaths = 574). There were 18 candidate variables (denoted here only by the numbers 1-18). Thirty resampling runs were done. The variables selected are shown in Table 1. The number of variables selected ranged from 3 to 10 (Figure 1). Some variables were selected only once, but no variable was selected every time. The purpose of this simulation is not to suggest removing stepwise regression from the repertoire of modeling techniques, but to urge caution in uncritically accepting the results.

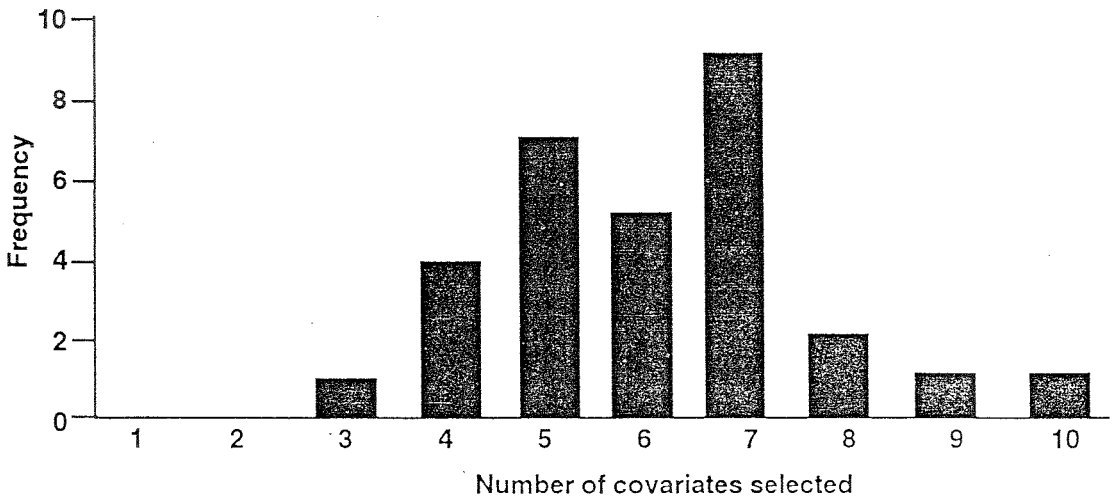
Several methods can be used to examine the reliability of the model and the reliability of the covariate estimates. For the overall model, one simple method is to plot the average model-predicted survival estimate against the corresponding Kaplan-Meier survival estimate for each decile of predicted survival. Estimates of survival can be analyzed similarly for subsets defined by covariates (for example, diabetics and nondiabetics), thus giving an estimate of how well the model predicts for a particular covariate. Note that in this latter case, the covariate defining the subset does not have to be a covariate included in the model. Sex may not affect survival, but a good model would predict well for both the male and female subsets. Because models developed on a particular data set generally estimate well within the same data set, however, additional methods should be employed.

One of these methods is data splitting. In its simplest form, the strategy is to divide the data set randomly into two subsets: a training data set and a test data set. The model developed on the training data set is then applied to the test set. Comparing the model predictions in the test data set with Kaplan-Meier estimates of survival will yield an estimate of model reliability.[1] Data splitting results in discarding a large fraction of the data for both model fitting and estimation of model reliability and, hence, is inefficient.

Data splitting can be improved by k-group cross-validation. This is done by subdividing the data set into k mutually exclusive and exhaustive subsets. The k separate models are developed using stepwise variable selection on the (k-1)/k fraction of the data, and the model is validated on the remaining 1/k fraction of the data. Es-

[1]This is also being done in a larger framework. There is considerable effort through the Chronic Ischemic Heart Disease Patient Outcomes Research Team (Pryor, D.B., principal investigator, AHCPR grant no. HS06353-01, Outcome assessment program in ischemic heart disease, 1989) to test models developed at one institution on data sets collected at other institutions.

Figure 1. Number of covariates selected in resampling simulation



Note: N = 984, deaths = 574

Source: Duke University Heart Center.

Table 1. Variables selected in a resampling simulation

	Variable																		
Run	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	n
1	•			•		•													3
2		•	•	•					•			•	•	•	•				8
3	•		•	•					•			•				•	•		7
4		•	•	•					•			•			•	•			7
5				•			•		•	•		•							5
6		•	•					•	•		•	•			•	•			8
7		•	•	•					•			•					•		6
8	•		•	•			•		•			•			•				7
9			•	•			•		•	•		•					•		7
10		•		•					•			•							4
11		•	•	•					•	•	•	•							7
12	•	•		•				•	•			•				•			7
13	•	•	•	•				•	•	•	•	•				•			10
14		•	•	•	•		•		•			•							7
15	•	•	•	•								•	•						6
16			•	•					•			•							4
17		•							•	•	•	•							5
18		•		•	•				•			•							5
19		•		•					•	•		•							5
20	•	•		•					•			•							5
21	•	•	•	•					•		•	•	•					•	9
22	•	•		•														•	4
23		•		•					•			•							4
24				•			•			•		•	•			•		•	7
25	•	•	•	•					•	•		•							7
26		•	•						•			•	•						5
27	•		•	•					•			•			•				6
28		•		•			•		•	•		•							6
29				•	•					•		•						•	5
30		•		•					•	•	•	•							6
N	11	21	16	27	3	1	6	3	25	11	6	28	5	1	5	6	3	4	

Run = Simulation run number
n = Number of variables selected for the run
N = Number of times a specific variable was selected

timates of fit (for example, R^2, difference between predicted and observed survival) are then averaged over the k validation runs. If the model validates, it is then fit to the entire data set for final parameter estimates. If the model does not validate, careful examination of the underlying hazard and candidate covariates should be undertaken.

Another method uses the bootstrapping technique (Efron, 1982). Here, the same set of data can be used to both fit and validate the model. Bootstrapping allows one to estimate and correct for the tendency of the predicted estimate to correlate better with Kaplan-Meier survival estimates than would be expected if the validation were done on separate data sets.

Indexes. A second consideration in variable selection is that several variables may jointly describe a particular aspect of a patient's condition, such as heart function or anginal pain. Because such an aspect of a patient's condition may manifest itself in different ways in different patients, the inclusion of all related variables in the model may dilute the overall effect of any individual component. Combining variables that describe the same condition into a composite index can yield more powerful results than using the individual variables. Harrell, Lee, Califf, and others (1984) gave an example in which 30 variables were reduced to 10 indexes. As an example, anginal symptoms in Table 2 form a composite based on three separate measures of anginal pain: severity, type, and frequency (Califf, Mark, Harrell, and others, 1988).

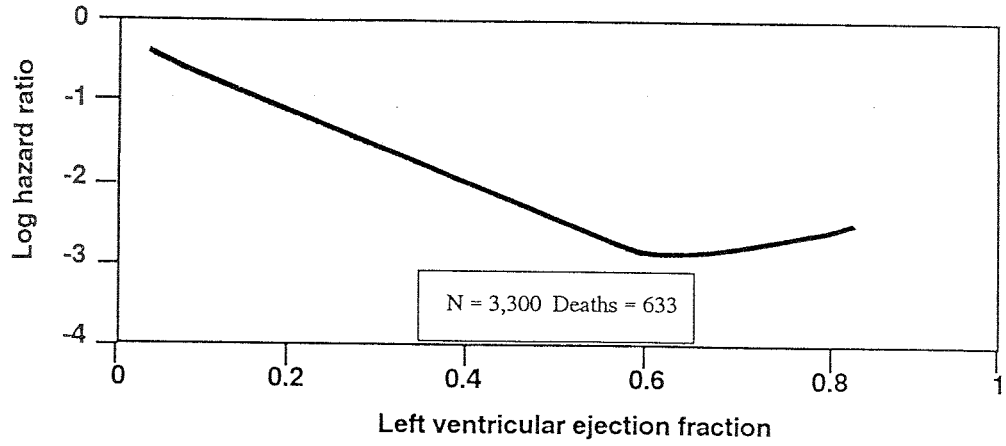
Functional form. The third problem in fitting a model is the functional form of the covariate. For example, is the incremental risk the same over the entire range of the covariate? Should x be used as measured or some function of x? Several investigators (Harrell and Lee, 1986; Harrell, Lee, and Pollock, 1988; Harrell, Pollock, and Lee, 1987; Stone and Koo, 1985; Durrelman and Simon, 1989; Sleeper and Harrington, 1990) described an approach in which splines are used to determine the shape of the function of the covariate. For instance, in the study by Califf, Harrell, Lee, and others (1989), the left ventricular ejection fraction (LVEF) for the population ranged from 6 to 91. A cubic spline fit to the data (Figure 2) indicated that an LVEF above 60 had no incremental benefit. By truncating the LVEF to 60, the model likelihood χ^2 was increased significantly. For more complex functional forms, each component of the spline fit is entered in the model. To test for an overall covariate effect, a χ^2 with k-1 degrees of freedom is computed where k is the number of knots (join points) required to model the effect. Linearity is tested with k-2 degrees of freedom.

Investigators (for example, Therneau, Grambasch, and Fleming, 1990; Schoenfeld, 1982; Pettitt and Daud, 1990) have proposed methodologies for examining the nature of the effects of covariables across time. The use of time-dependent covariates to model the covariables' effect over time is also possible with parametric survival models (Herndon and Harrell, 1989).

Number of covariates. The fourth problem is the number of covariates a given data set can support. It is well known—but seldom taken into account—that as the number of parameters estimated exceeds a given value, predictions on a new sample actually become worse. This problem is particularly acute with regard to estimation of absolute survival probabilities.

The most reliable predictor is one that always predicts the mean outcome. If the end users desire discrimination ability, some reliability has to be sacrificed. The question then becomes "How many parameters can one estimate and still have reasonable reliability?" In fact, if

Figure 2. Spline transformation for left ventricular ejection fraction



Source: Duke University Heart Center.

too many parameters are estimated, even the discrimination will worsen due to overfitting (Harrell, Lee, Califf, and others, 1984).

To demonstrate this effect, we did two simulation experiments involving fitting Cox models with p uniformly distributed (on [-0.5,0.5]) covariates, x, when p is varied from 1 to 29. In the first experiment, for each p, 10 samples of size n = 750 were drawn from an exponential distribution with true hazard $\lambda\exp(\underline{x\beta})$ where $\underline{x\beta}=x_1\beta_1+x_2\beta_2+\cdots+x_p\beta_p$, and $\beta_1=\beta_2=\cdots=\beta_p=0.25$. λ was chosen to yield an average 5-year survival of 0.75. The samples had an average of 185 deaths. For each sample, a Cox model was fitted to the p covariates, and the predicted 5-year survival was computed for all subjects. These estimates were compared with the true survival probabilities for the same subjects. Average absolute errors were averaged over the 750 subjects and over the 10 samples. A second experiment was done with n = 250, with an average of 64 deaths at 5 years.

In Figure 3, the solid lines represent results from the n = 750n experiment and the dashed lines represent results from the n = 250 experiment. Results for the "sick" (high-risk) patients are denoted with boxes (■). Lines without boxes denote results for the "average" patient. The curves for the average patients were constructed by averaging the error over the entire patient population. The curves for the sick patients were constructed by averaging over patients from the lower decile of predicted survival.

If the investigator wishes to limit the average error to some fixed value, say 0.05, there is an upper limit on the number of parameters to be estimated. For the average patient, in the n = 750 experiment, the expected survival error is monotonically increasing in p and exceeds 0.05 when p > 14. For the n = 250 experiment, the error reaches 0.05 when p > 5. Absolute survival error rises more steeply for the highest decile of predicted risk. Survival estimation is often crucial for these sickest patients. Here, the error exceeds 0.05 for p > 5 (n = 750) and for p > 3 (n = 250). Thus, when the number of parameters is large compared with the number of events in the sample, overfitting occurs. This overfitting has more effect for patients with very high or very low predicted survival, because overfitting causes predicted survival to shrink toward the mean survival. Harrell, Lee, Matchar, and Reichert (1985) proposed that there should be at least 10 events for every candidate covariate in order to guard against this overfitting. Other work has shown this applies almost equally in a stepwise variable selection where p is the number of candidate variables. We need to constantly remind ourselves of the penalty for creating very complex models.

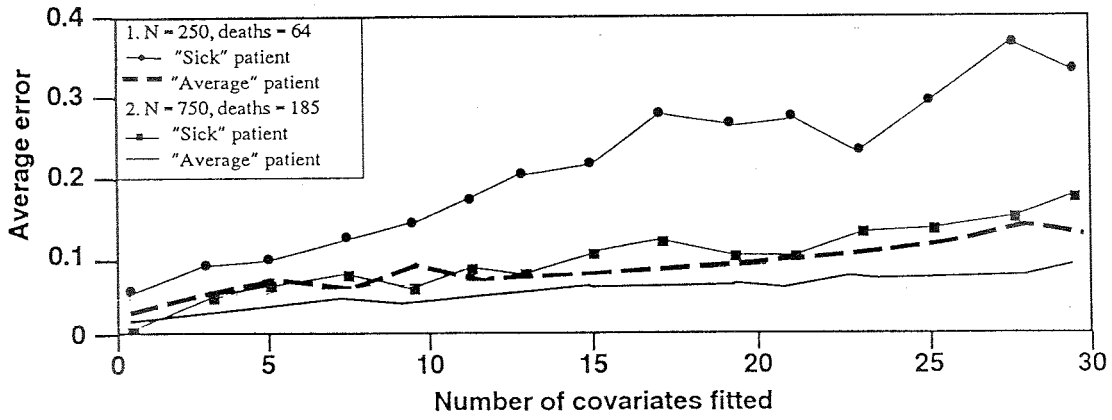
One final point is that covariate selection should be done with the physician and the statistician working as a team, each supplying their own particular expertise and knowledge to the enterprise.

Prediction

Broadly speaking, there are two functions of prediction—forecasting and hypothesis testing (see Bunge, 1979, for a general discussion). In survival analysis, forecasting is the deduction of future probabilities of survival for a given patient or population. An example of hypothesis testing is confirming or refuting the hypothesis that a specific covariate influences survival.

Naftel (in this volume) points out that survival models provide curves for each treatment to both the patient and the physician; these curves represent the most likely probability of survival at each time point. Survival prediction serves a number of worthwhile purposes.

Figure 3. Average error in Cox 5-year survival estimates



Source: Duke University Heart Center.

By comparing the curves over time, the patient and physician can make a more informed decision about treatment selection. If the patient's personal goal is to maximize the likelihood of short-term survival, a specific treatment may be chosen. If the patient's goal is to maximize long-term survival, an alternative treatment may be chosen.

On a larger scale, the prediction curves can provide information to public and private health care providers about the most likely survival curve for specific patient groups. Given the predicted survival curves, these agencies may recommend different treatments and reimbursement strategies for different patient groups.

Multiphase Model

The multiphase model is structurally more complex than other parametric models. The primary goal in its development was to model a wide variety of survival shapes and provide a means to determine how covariates influence various parts of the survival curve.

Strengths. The multiphase model can estimate the effects of covariates at different time segments along the survival curve. Table 2 lists covariates that influenced outcome in 2,967 coronary artery bypass graft patients (Califf, Harrell, Lee, and others, 1989). Smith, Harrell, Rankin, and others (1991) used the multiphase model to analyze these data. Table 2 lists these results along with those from an analysis using logistic regression for 30-day mortality and with results from the Cox model for long-term survival. A comparison of these three analyses shows a strength of the multiphase model.

The early phase mimics results of the logistic model. Both have the ability to estimate covariate effects when the covariate is influential for a short time after treatment is begun. It is possible that the initial hazard due to the treatment may last for shorter or longer periods of time depending on other patient characteristics. In such a case, the logistic model using fixed-length followup (in-hospital, 30-day, or 6-month mortality) may not be adequate to describe accurately the relevant covariates that affect the early treatment hazard. The multiphase model can estimate such effects more accurately, provided that patients are followed long enough to define the early phase accurately.

The Cox model identifies covariates that are found in at least one of the phases of the multiphase model. The Cox model "averages" over the entire hazard, whereas the multiphase model identifies covariates that dominate the hazard at particular times. Thus, the Cox model and the multiphase model tend to give consistent results overall, but the multiphase model yields additional information on when the covariates most influence survival. A modification of the Cox model can permit the partition of survival into separate phases but at the expense of considerable computational burden (Anderson and Senthilselvan, 1982).

Another strength of the model is that it is implemented as a procedure in SAS (Blackstone and Naftel, 1988), one of the most versatile, powerful, and widely used of the extant statistical analysis software packages. Thus, data input and results output are relatively straightforward.

Weaknesses. The learning curve for using this model is steep and daunting. In order to model some survival curves accurately, up to three different hazard functions must be specified with as many as nine shaping parameters (four in the early phase, one in the middle phase, four in the late phase) to be estimated. One successful strategy for estimating the hazard parameters is to include only one distribution and achieve the best fit, then fix the parameters and add a second phase; when the best fit is achieved, fix the parameters in the second phase and add the third phase. When parameters for this last phase have been estimated, constraints on all parameters are removed and the model is refitted to get a complete set of hazard parameter estimates. At each step, the significance of the phase can be assessed with the likelihood ratio test.

When all the phases have been identified, the hazard parameters are fixed, and the separate streams of covariates are added. When covariate selection is completed, all constraints on the hazard parameters are released and the model is run again to get final estimates for the hazard parameters, covariate effects, and the covariance matrix. This entire procedure can be long, onerous, and

Table 2. Comparison of covariates selected by three survival models

Covariate	Logistic 30 day death	Multiphase[1] early	Multiphase[1] late	Cox model
Earlier year of surgery	•	•	•	•
Lower ejection fraction	•	•	•	•
Greater extent of coronary artery disease	•	•	•	•
Older age	•	•	•	•
More conduction abnormality	•	•	•	•
History of hypertension	•	•	•	•
Greater anginal symptoms	•	•		•
Lower weight	•	•		•
Greater vascular disease			•	•
Presence of diabetes			•	•
Greater myocardial damage			•	•

[1]Multiphase analysis originally published by Smith, Harrell, Rankin, and others, 1991.

Note: A dot (•) indicates that covariate was significant ($p < .05$).

expensive. In addition, up to three streams of covariate effects may be modeled. Thus, the previously described problems of finding spuriously significant covariates are compounded greatly.

As currently implemented, there is no ability to test for equal covariate effects across phases during the stepwise covariate selection process. Only after covariate selection can a test be done on those covariates selected for two or more phases. If a covariate is selected for one phase but not for another, no test of equality of effect is possible unless the covariate is forced into the model. Furthermore, if a covariate has a similar effect over two or more phases, there is no facility for estimating a common effect. The next version of the model, however, will have this very useful facility (Naftel, personal communication, September 1991).

We recommend that the papers by Blackstone, Naftel, and Turner (1986), Bradley, Bradley, and Naftel (1984), Turner (1975), Turner, Hazelrig, and Blackstone (1982), and Turner and Pruitt (1978) be reviewed in order to understand the interrelationships among the distribution subfamilies.

Summary

Many models are available for use in assessing treatment effects, including the nonparametric models such as the Cox model, the parametric variations on the Cox model (Herndon and Harrell, 1989), and other parametric models (for instance, Bailey, 1988), as well as the multiphase model. These models may incorporate separate streams of covariates or use time-dependent covariates to assess treatment effect. Each of these models is under continual development and offers great promise for more accurate identification of important covariates.

No model or process completely solves all the problems outlined here, but the multiphase model is a new and powerful analytic technique for treatment comparison, covariate selection, and prediction and deserves further evaluation. It has been used successfully in a wide range of applications requiring many different hazard shapes.[2] The hallmark of any survival modeling strategy should be simplicity, and everyone engaged in this pursuit should carefully examine whether the problem and the quality and quantity of data at hand warrant the complexity of the model (or indeed, any model) being used. Also, the more complex the model, the more important validation becomes.

[2]The American College of Cardiology/American Heart Association Task Force Report (1991) has extensive examples of its use in studies of coronary artery disease, and an extensive bibliography.

[3]Pryor, D.B., Principal investigator, AHCPR grant no. HS06353-01, Outcome assessment program in ischemic heart disease, 1989.

Other issues that were not discussed by Naftel (in this volume) but that ultimately bear on the survival analysis problem include the use of treatment data not collected in a randomized clinical trial and the analysis of treatment data where treatment is confounded with data collection site. These issues have been addressed (Hlatky, Califf, Harrell, and others, 1988)[3], and continue to be studied.

The prediction process has some interesting ethical ramifications. The method permits an objective assessment of risk. The patient and the health care provider, however, looking at the same treatment survival curves may make two diametrically opposite choices. Conflicts may arise if the health care providers are unwilling to pay for the patient's choice. Thus, there is potential for conflict between two sets of ethics: those of the patient as an individual and those of society as a whole.

Survival analysis, with its ability to predict the benefit of one treatment over another, is only one part of the larger problem of how to allocate health care resources. The next step is to be able to predict the cost (either cost to the patient, cost to the health care provider, or in some sense, cost to society) with the same precision and accuracy that we can predict benefit. The modeling process could play a central role in the resolution of this important problem.

References

American College of Cardiology/American Heart Association Task Force Report. (1991). Guidelines and indications for coronary artery bypass graft surgery. *Journal of the American College of Cardiology*, 17(3), 543-589.

Anderson, J.A. and Senthilselvan, A. (1982). A two-step regression model for hazard functions. *Applied Statistics*, 31(1), 44-51.

Bailey, R.C. (1988). Some uses of a modified Makeham model to evaluate medical practice. *Journal of the Washington Academy of Sciences*, 78(4), 339-353.

Blackstone, E.H. and Naftel, D.C. (1988). *SAS technical report P-175, Changes and enhancements to the SAS system, Release 5.18, under OS and CMS*. Cary, NC: SAS Institute.

Blackstone, E.H., Naftel, D.C., and Turner, M.E., Jr. (1986). The decomposition of time-varying hazard into phases, each incorporating a separate stream of concomitant information. *Journal of the American Statistical Association*, 81(395), 615-624.

Bradley, D.H., Bradley, E.L., and Naftel, D.C. (1984). A generalized Gompertz-Rayleigh model as a survival distribution. *Mathematical Biosciences*, 70(2), 195-202.

Bunge, M. (1979). *Causality and modern science* (3rd rev.) (pp. 310). New York: Dover Publications.

Califf, R.M., Harrell, F.E., Jr., Lee, K.L., and others. (1989). The evolution of medical and surgical therapy for coronary artery disease: A 15 year perspective. *Journal of the American Medical Association*, 261, 2077-2086.

Califf, R.M., Mark, D.B., Harrell, F.E., Jr., and others. (1988). Importance of clinical measures of ischemia in the prognosis of patients with documented coronary artery disease. *Journal of the American College of Cardiology*, 11(1), 20-26.

Cox, D.R. (1972). Regression models and life-tables (with discussion). *Journal of the Royal Statistical Society*, Ser B, 34(2), 187-220.

Durrelman, S. and Simon, R. (1989). Flexible regression models with cubic splines. *Statistics in Medicine*, 8, 551-561.

Efron, B. (1982). The jackknife, the bootstrap, and other resampling plans. *CBMS-NSF regional conference series in applied mathematics*, Vol. 38. Philadelphia: Society for Industrial and Applied Mathematics.

Efron, B. and Gong, G. (1983). A leisurely look at the bootstrap, the jackknife, and cross-validation. *The American Statistician*, 37(1), 36-48.

Feigl, P. and Zelen, M. (1965). Estimation of exponential survival probabilities with concomitant information. *Biometrics*, 21, 826-838.

Harrell, F.E., Jr. and Lee, K.L. (1986). Verifying assumptions of the Cox proportional hazards model. *Proceedings of the 11th Annual SAS User's Group International Conference* (pp. 823-828). Cary, NC: SAS Institute.

Harrell, F.E., Jr., Lee, K.L., Califf, R.M., and others. (1984). Regression modeling strategies for improved diagnostic prediction. *Statistics in Medicine*, 3, 143-152.

Harrell, F.E., Jr., Lee, K.L., Matchar, D.B., and Reichert, T.A. (1985). Regression models for prognostic prediction: Advantages, problems, and suggested solutions. *Cancer Treatment Reports*, 69(10), 1071-1077.

Harrell, F.E., Jr., Lee, K.L., and Pollock, B.G. (1988). Regression models in clinical studies: Determining relationships between predictors and response. *Journal of the National Cancer Institute*, 80(15), 10-13.

Harrell, F.E., Jr., Pollock, B.G., and Lee, K.L. (1987). Graphical methods in survival analysis. Proceedings of the 12th Annual *SAS User's Group International Conference* (pp. 1107-1115). Cary, NC: SAS Institute.

Herndon, J.E., II and Harrell, F.E., Jr. (1989). Examining the relationship between survival time and time dependent covariables using the PHSPLM procedure. *Proceedings of the 14th Annual SAS Users Group International Conference* (pp. 1318-1323). Cary, NC: SAS Institute.

Hlatky, M.A., Califf, R.M., Harrell, F.E., Jr., and others. (1988). Comparison of predictions based on observational data with the results of randomized controlled clinical trials of coronary artery bypass surgery. *Journal of the American College of Cardiology*, 11(2), 237-245.

Pettitt, A.N. and Daud, I.B. (1990). Investigating time dependence in Cox's proportional hazards model. *Applied Statistics*, 39(3), 313-329.

Schoenfeld, D. (1982). Partial residuals for the proportional hazards regression model. *Biometrika*, 69(1), 239-241.

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

Smith, L.R., Harrell, F.E., Jr., Rankin, J.S., and others. (1991). Determinants of early versus late cardiac death in patients undergoing coronary artery bypass graft surgery. *Circulation*, 84(Suppl. III), 245-250.

Stone, C.J. and Koo, C.Y. (1985). Additive splines in statistics. *Proceedings of the Statistical Computing Section* (pp. 45-48). Washington: American Statistical Association.

Thall, P.F. and Lachin, J.M. (1986). Assessment of stratum-covariate interactions in Cox's proportional hazards regression model. *Statistics in Medicine*, 5, 73-83.

Therneau, T.M., Grambasch, P.M., and Fleming, T.R. (1990). Martingale-based residuals for survival analysis. *Biometrika*, 77(1), 147-160.

Turner, M.E., Jr. (1975). Some classes of hit-theory models. *Mathematical Biosciences*, 23, 219-235.

Turner, M.E., Jr., Hazelrig, J.B., and Blackstone, E.H. (1982). Bounded survival. *Mathematical Biosciences*, 59, 33-46.

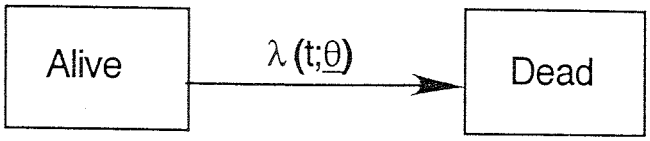
Turner, M.E., Jr. and Pruitt, K.M. (1978). A common basis for survival, growth, and autocatalysis. *Mathematical Biosciences*, 39, 113-123.

Appendix

Survival models. In general, survival models can be thought of as a special case of a two-state (alive, dead) Markov process (appendix figure). The transition rate from the alive state to the dead state is governed by the hazard function $\lambda(t;\underline{\Theta})$ where t is time and $\underline{\Theta}$ is a vector of parameters that govern the shape of the hazard function. The corresponding survival function is $S(t)=\exp\{-\Lambda(t,\underline{\Theta})\}$, where $\Lambda(t,\underline{\Theta})$ is the cumulative hazard up to time t. The central issue of survival analysis is to describe $\lambda(t;\underline{\Theta})$ and to determine what covariates modify $\lambda(t;\underline{\Theta})$. There are two general methods of incorporating covariates into the model. The first is to assume that one or more of the parameters, Θ, are linear combinations of the covariates, $\underline{x}$ (eg $\Theta_j=\underline{x}\beta$). This

method was used early by Feigl and Zelen (1965) and more recently by Bailey (1988).

Appendix figure.



Another method is to multiplicatively scale $\lambda(t;\underline{\Theta})$ by $\exp(\underline{x}\underline{\beta})$ so that the adjusted hazard is $\lambda(t;\underline{x},\underline{\Theta}) = \lambda(t;\underline{\Theta}) \exp(\underline{x}\underline{\beta})$. This is the proportional hazards formulation. If a specific parametric form of the hazard function is assumed, the model is called parametric. If no specific function form is assumed, then the model is semi-parametric (Cox, 1972).

The multiphase model assumes the cumulative hazard to be the weighted sum of three hazards, where the weights are scaling functions of the covariates. One phase describes the early phase of survival. A second phase describes intermediate survival. The third phase describes late survival. Each of these three phases is present over all time but dominates survival during only one time interval. The total cumulative hazard function is the sum of the three phases; that is

$$\Lambda_T(t) = \mu_1(\underline{x}_1,\underline{\beta}_1) \cdot \Lambda_1(t,\underline{\Theta}_1) + \mu_2(\underline{x}_2,\underline{\beta}_2) \cdot \Lambda_2(t,\underline{\Theta}_2) + \mu_3(\underline{x}_3,\underline{\beta}_3) \cdot \Lambda_3(t,\underline{\Theta}_3), \Theta$$

where, for i=1(early), 2(middle), and 3(late), $\mu_i=\exp(\underline{x}_i\underline{\beta}_i)$ is the scaling parametric function of the covariates, $\underline{x}_i$=vector of covariates, $\underline{\beta}_i$=vector of regression coefficients, Λ_i=cumulative hazard function, $\underline{\Theta}_i$=vector of parameters governing the shape of the cumulative hazard, and t=time to failure. The specific functional forms of Λ_i are given by Blackstone, Naftel, and Turner (1986). The survival function is given by $S(t)=\exp\{-\Lambda_T(t)\}$. The model permits the description of a survival function using any one, two, or all three of these phases.

The form of the scaling parametric functions, μ_i, is equivalent to a proportional hazards assumption for each phase of the model. Because the model can describe each phase of the survival curve by a different set of covariates, the separate phases of the survival curve can be adjusted independently.