A SURVEY OF MICROCOMPUTER SURVIVAL ANALYSIS SOFTWARE: THE NEED FOR AN INTEGRATED FRAMEWORK

Frank E. Harrell, Jr. Richard Goldstein

April 3, 1997

ABSTRACT

This paper surveys commercially available MS–DOS and Microsoft Windows based microcomputer software for survival analysis, especially for Cox proportional hazards regression and parametric survival models. Emphasis is given to functionality, documentation, generality, and flexibility of software. A discussion of the need for software integration is given, which leads to the conclusion that survival analysis software not closely tied to a well–designed package will not meet an analyst's general needs. Some stand–alone programs are good tools for teaching the theory of some survival analysis procedures, but they may not teach the student good data analysis techniques such as critically examining regression assumptions. We contrast typical software with a general, integrated, modeling framework that is available with S-PLUS .

Key words: survival analysis, software, microcomputers, parametric survival models, Cox proportional hazards model

---

From the Division of Biostatistics and Epidemiology, Department of Health Evaluations Sciences, School of Medicine, University of Virginia (FH) and Qualitas Inc., Brighton MA (RG)

This work was supported by Research Grants HS–03834, HS–04873, HS–05635, HS–06503, HSS–06830, and HS–07137 from the Agency for Health Care Policy and Research, Research Grant HL–17670 from the National Heart, Lung, and Blood Institute, and grants from the Robert Wood Johnson Foundation, Princeton NJ, and the National Center for Health Statistics.

A very early version of this paper appeared in the *Proceedings of the Section on Statistical Education ASA, 1988.*

Address for correspondence and reprint requests:

Frank E Harrell, Jr, PhD
Box 600 Health Sciences Center
University of Virginia
Charlottesville VA 22908
E–mail: fharrell@virginia.edu

1 INTRODUCTION

For almost 20 years mainframe computer software has been widely available for survival analysis. SAS and BMDP have led the way in implementing reliable parametric and semi–parametric regression models and other techniques for analyzing censored failure time data. While general statistical software for microcomputers started to become available a few years ago, the implementation of survival analysis procedures has lagged behind. Recently, however, a number of promising statistical packages and stand–alone programs have begun to be marketed. This software addresses some of the most important survival analysis needs. This paper surveys the capabilities of some MS–DOS/Windows software and describes the software's flexibility as well as the adequacy of documentation.

Some of the software reviewed is intended to be self–contained. Examples are given in which this philosophy places significant restrictions on the analyst.

See Goldstein, Anderson, Ash, Craig, Harrington & Pagano (1989) for a comprehensive review of several microcomputer survival analysis programs. Other reviews of survival analysis software not limited to the PC are found in Wagner & Meeker (1985), Hudec, Hitz & Muellner (1985), and Dain, Freeman & Vredenburgh (1989). Also, see Huster & Enterline (1985) for a description of an interactive package of basic survival analysis methods suitable for the classroom, and Iwane, Palensky & Plante (1997) for a review of software for sample size calculations for survival analysis.

2 PURPOSES OF MICROCOMPUTER STATISTICAL SOFTWARE

Depending on the size of the dataset being analyzed, microcomputers can be efficient and cost–effective tools for data analysis. Microcomputers are also being increasingly used in teaching statistics. Listed below are three categories of use for survival analysis software, the last two both involving teaching.

1. Routine data analysis.
2. Teaching isolated components of the statistical theory involved with a given statistical tool. Examples: understanding distributional shapes and interpretation of parameter estimates, standard errors, and various test statistics.
3. Teaching how a given tool is used as part of a complete data analysis strategy. Examples: testing linearity of a predictor variable in a Weibull regression model, graphically and analytically estimating a nonlinear transformation to make the predictor satisfy regression assumptions, fitting a general transformation using the standard maximum likelihood technique, and testing for additivity (lack of interaction).

In the survey, attention is paid to these three categories of use.

3 SURVIVAL ANALYSIS METHODS REVIEWED

There are many statistical methods useful for analyzing survival time data. The methods reviewed in this paper are parametric survival models, with and without covariables (Lawless 1982), and the Cox (1972) semiparametric proportional hazards regression model (CPH). We do not give emphasis to Kaplan & Meier (1958) (KM) estimation, statistical tests of survival distributions, and specialized procedures for the following reasons:

1. There is little difference between the packages in KM except for the computation of confidence intervals;
2. regarding statistical tests, there is controversy regarding both
   (a) whether they should be used (frequently a statistical test is a special case of a more general regression model, e.g. the log–rank test vs. score statistics from CPH), and
   (b) what is meant by each test (Dain et al. 1989, Goldstein et al. 1989); that is, different vendors seem to use the same name for formulas that are similar but not quite the same;
3. only one of the included packages deals with specialized setups such as frailty models (unobserved heterogeneity); and
4. all three issues were discussed in Goldstein et al. (1989) and there has been little or no change in these areas in the intervening years.

The majority of packages allow the covariables to have time–dependent effects or to vary over time in the CPH model. In what follows we will use the term *time–varying covariable* to refer to a variable whose values are updated over time, forming a step–function of time, and we use the term *time–dependent covariable* to refer to an interaction between a baseline (constant) covariable and some function of time. See Goldstein et al. (1989) for more about the distinction of the two types of time effects, and Altman & De Stavola (1994) for useful advice on conducting analyses of time–varying covariables.

We briefly reviewed the packages for some advanced survival analysis capabilities which are discussed separately (Section 8).

4 SOFTWARE CAPABILITIES REVIEWED

Within the set of methods listed in the last section, we examine both the technical options available, such as what types of models are allowed and what types of time–varying or time–dependent covariables are allowed, and the user interface for using each of these, such as how one sets up the data set, how easy it is to replicate an analysis, and what output from an analysis can be saved into a data set for further processing. For the most part, we do not mention particular algorithms or techniques used in estimation unless some problem is noted. Similarly, several options that we consider necessary are not mentioned unless either there is something odd in the way the software provides them or the software does not provide them at all. For example, the likelihood ratio (LR) test option is crucial as it provides a simultaneous test of all covariables, adjusting for multiple comparisons. This test will help the analyst guard against examining individual "significant" effects in a "fishing expedition" (and thus overfitting the data) when there is insufficient evidence to warrant such pursuits. SURVIVAL and SAS procedure `LIFEREG`, for example, do not print the LR test statistic even though they print half of this statistic — the maximized log–likelihood for the model. EGRET prints the correct LR statistic for CPH, but its LR tests for other models containing an intercept are sometimes inappropriate — users must become familiar with the interactive modeling style of EGRET by remembering to fit the `%GM` (grand mean) term first so that the first reported LR statistic will be appropriate.

The ability to accept user–specified covariable and strata values for obtaining predictions, and to store predictions in a standard–format dataset are vital for routine data analysis as well as for model verification and display. It is beneficial for the results from the model fit to be stored in a way that is easily accessed by other programs in the environment in which the analysis program exists. For example, these features can be used to obtain predictions on an independent sample not used in fitting the model, without the user having to type the coefficients of the fitted model. Prediction output options can be used to send predicted survival and hazard curves to a general–purpose graphics program to display results of the final analysis, and to compute confidence limits when only standard errors are available. Examples of the importance of these capabilities are given in Harrell (1983), which concludes that without these capabilities, survival software will be unable to perform these commonly needed analyses:

1. Use a stratified CPH model to plot "adjusted KM" survival curves, with modeled covariables adjusted to a computed value such as the overall median.

2. Display the effect on survival of a continuous covariable by plotting 2– and 5–year survival over a continuous range of the covariable.

3. Automatically stratify a continuous variable into intervals containing at least 40 observations. Display the KM 2– and 5–year survival estimates vs. the mean covariable value within each interval. These empirical estimates are compared with predictions from a model, with the mean value within each interval used to obtained predictions that should approximately match the KM estimates.

These types of input and output options are used in the SAS macro procedure `SRVTREND` [1] (Harrell, Pollock & Lee 1987), in which the user can display survival estimates with respect to 1, 2, or 3 varying factors, each with its own level of assumption in the context of a CPH model: no assumption (stratify, with automatic formation of quantile groups for continuous factors), proportional hazards without assuming linearity (by automatically fitting a restricted spline function in the covariable to estimate the transformation that would make the covariable fit the model linearly (Harrell, Lee & Pollock 1988)), and proportional hazards assuming linearity. Any number of additional covariables may be adjusted for but not displayed graphically. All transformations, design matrices, and needed SAS code are generated automatically by `SRVTREND`. Figure 1 displays the relationship between two continuous variables, `X1` and `X2`, and 5–year survival probability derived from a sample

[1] To obtain `SRVTREND` and related SAS macros, send E–mail to the software archive operated by the Journal of Statistical Education, North Carolina State University, `archive@jse.stat.ncsu.edu`. Include the message `send software/sas/survrisk.pak`.

of size 3000. Proportional hazards but not linearity is assumed for `X1` and nothing is assumed for `X2`. Estimates are adjusted to the grand mean for `X3`, a variable which is not displayed for which proportional hazards with linearity is assumed. That is, a 5–knot restricted cubic spline function (with 3 parameters) is fitted with respect to `X1`, a linear term is fitted for `X3`, and quartile groups are formed automatically for `X2`, with groups identified on the graph by the mean `X2` in each quartile group. A CPH model stratified by `X2` quartile and with 5 covariables (`X1`, `X3`, and 3 spline terms for `X1`) is used to derive these estimates. The figure was produced with the single SAS statement

```
%SRVTREND(TIME=t,EVENT=e, X=x1, XASSUME=2, NK=5,
          RANGE=0 TO 1 BY .02, CLASS=x2, CGROUPS=4,
          ADJ=x3, T=5, PLOT=2)
```

Much more general analyses can be performed using S-PLUS (Becker, Chambers & Wilks 1988, MathSoft 1995), as will be discussed in section 9.

[**Figure 1 about here**]

5 STRUCTURE OF REVIEWS

The following items are discussed regarding each package:

1. What type of package is it: general purpose with survival module, specialized survival package, etc.
2. User interface: menu, command driven, interactive; can one extend the package via either macros or a language; etc.
3. What models are present in the package; here, something is only considered to be present if it is discussed in the documentation, whether that means the main manual, a supplementary manual or the on–line help.
4. Ease of Learning and Ease of Use of the Package, with particular emphasis on going beyond what the package expects the user to do; for example, can output be saved and either used or exported to another packages for use? Note that ease of use refers to use by someone who is not a beginner with that package.
5. Data set limitations, if any.
6. Assessment of the usefulness of the software in the three settings discussed in Section 2, above.

It is very important in using statistical software that the package allow the user to obtain a complete audit trail of all instructions given to the package by the user (Goldstein 1991). All included packages, including the menu–driven packages, EGRET and TRUE EPISTAT, have such a feature.

Two of the packages, `EGRET` and `EPILOG PLUS`, are copy protected.

6 REVIEWS OF INDIVIDUAL PACKAGES

Letters were originally sent to vendors of MS–DOS software in the winter of 1987–1988. Not all vendors responded; further, software of some respondents is not included here. For example, SURV-CALC has previously been shown to have several bugs (Clayton 1988, Goldstein et al. 1989) which have never been fixed. On the other hand, some vendors have subsequently sent material which has been included. All included software is listed in the appendix with addresses and phone numbers of vendors. Summaries of the reviews, based both on reading the documentation and using the software follow, the order of the packages determined alphabetically by the name of the package. Not reviewed are GAUSS, which has some undocumented user–written functions, and MINITAB, which has a user–contributed macro for unstratified KM.

6.1 BMDP Version 7.0

BMDP is a command–oriented general purpose package that can be run either interactively or in batch mode and has some menu capabilities. On–line help is available. The PC version of the package cannot be extended, though one can write macros to make the use of frequently used instructions more convenient.

BMDP includes both Cox's semiparametric proportional hazards model and accelerated failure time parametric models. The user can choose between the Exponential, Weibull, log–logistic and log–normal distributions for the parametric models. One module, BMDP2L, does competing risk analysis.

A flexible transformation language makes available to the user both an algebraic processor and "if–then" language allowing specification of a wide range of types of time–dependent and time–varying covariables (see Hoffman (1985) for examples of using BMDP for time–dependent covariables). In addition, both standardized and Cox–Snell residuals are explained and can be used in plotting; they can also be output. However, there is no way to obtain predictions for covariable settings stored in a data set or for outputting survival estimates for further processing.

The manual provides the best description of the underlying assumptions and how to test them of all manuals considered; BMDP is the only program that discusses the log–linear assumption and how to test it. Unfortunately, the manual is not complete regarding the survival programs and should be supplemented with an excellent technical report (No. 80), especially if the user has data with changing risk groups and thus needs to control what cases are in the risk group at each point in time; this important capability is present in the program but RG only found it usable via the technical report—neither the manual nor technical support were any help. The program is able to perform joint significance tests on combinations of coefficients using either Wald, LR, or score tests. Stepwise regression is available, or variables can be selected interactively by the user. Time–to–response data can be entered either as one variable giving this information or as a set of calendar date variables; the program will then calculate the time–to–response.

The program is relatively easy to learn and to use—certainly much easier than it was even a few years ago. However, if you don't use the program for several weeks, you are likely to find that you must re–learn the program, as the command structure is not intuitive. The large number of complete examples in the manual partly compensates for this, as does the new customizable editor. The high–resolution graphics have also been improved. Many valuable graphs are routinely available, and discussed in the manual, including a graph of the censoring pattern (also available in SURVIVAL and S-PLUS).

BMDP is an excellent tool for routine data analysis. With the addition of a version that uses expanded memory, the program can now operate on quite reasonably–sized data sets. The clear examples and explanations in the manual make it a very useful teaching tool also. However, the lack of integration of the various BMDP modules makes integrating the data analysis harder than it should be.

6.2 EGRET, Version 1.02.10

EGRET is a menu–driven program specialized for the use of "regression" analysis in medical studies. This includes: Cox models; parametric models including Exponential and Weibull; Poisson, logistic and conditional logistic models; and, random effects logistic regression. A generalization of the Weibull model in which the scale parameter can vary with covariables is also implemented (this is related to Stata's Bailey–Makeham model). Each model is discussed at several different levels of detail, from the theoretical mathematical to brief English overviews. Help is even provided in choosing between various models. The package cannot be extended by the user.

A strong feature of EGRET is that it allows the user to specify one of three iterative estimation algorithms and to select between an approximate and an exact analysis (which fully handles tied failure times).

EGRET is easy to learn and relatively easy to use, especially if one uses the limited macro capability. A unique feature of EGRET is that a default analysis model is stored with each dataset. This feature speeds and simplifies later analyses. Context–sensitive on–line help is available and one can annotate the output file "on the fly". Dummy variables are automatically generated (although classification variables must be integer valued), and the package offers the user the option to provide a joint test of significance of the set. The manual contains a clear discussion of testing the proportional hazards assumption, though no mention is made of the log–linear assumption. EGRET is one of the packages, with BMDP, Epicure, SAS, S-PLUS , and Stata, to offer the user the ability to control which subjects are in the risk group at any given time. This is particularly important when subjects are in competition; e.g., when studying promotion or when studying who gets a heart transplant.

Both continuous and step–function time–varying covariables are allowed. Up to 20 time–varying covariables can be defined at any one time; every subject can have different change points, and up to twenty pairs of time–change/new value variables can be defined for any one model.

Predicted values and delta–betas are available. Computed values can be graphed or can be listed or saved to the output file. The high resolution graphics option is particularly valuable, allowing the user to choose such things as: line styles that can differ for cases and controls or for censored and non–censored observations; and, labeling selected observations. Data sets are limited to 9,000 observations when there are time–varying covariables in the Cox model.

Data management is the clear weak point of this package, including very limited transformation capabilities. This is somewhat eased by the use of DBMS/Copy (Hallahan 1992) for moving files between EGRET and other packages, though this does not help much with regard to output data such as predicted probabilities. Further, the data manipulation capabilities and the estimation capabilities are in different, non–integrated programs which RG found quite awkward.

The package is useful for data analysis and for teaching, though the data management weaknesses affect both types of use. The excellent multi–tiered discussion of the various models makes it particularly useful for teaching these models in a stand–alone fashion. A significant asset is that the package comes with an encyclopedia of datasets useful for exploring the program as well as providing instruction in the statistical methods used. The datasets are tied to references of published analyses by major contributors in the field.

6.3 Epicure 2.0

Epicure is a set of five interactive command-driven programs for risk regression. Of these programs, three are relevant here: `PEANUTS`, for Cox regression using individual data; `AMFIT`, for survival analysis of grouped data via Poisson regression; `DATAB`, for tabular analysis including the tables for input to `AMFIT`. There are no parametric models for individual data, though one can use `AMFIT` for standard Poisson models. Both Kaplan–Meier and Nelson–Aalen nonparametric estimates are available.

Data manipulations, including transformations, are available in each module. Indicator variables and two-way interactions are allowed when entering a model estimation command. Graphs are low resolution, but the underlying data can be saved for export to other programs.

A unique feature in both `PEANUTS` and `AMFIT` is that additive and mixed multiplicative-additive models can be fit in addition to the standard multiplicative (relative risk) models. Both stratification and BY group processing are allowed. Both time–dependent and time–varying covariables may be used. Specification of models using the counting process methods are available, as is delayed entry into risk groups. Score and likelihood ratio tests, and confidence intervals, are available, in addition to standard Wald tests and intervals. Several types of residuals are available in both `PEANUTS` and `AMFIT`, as are delta-betas, delta-deviances, and some other diagnostics. The package is not extendible by the user. However, numerous derived variables can be output and saved or graphed. Thus, Figure 1, or at least the data underlying Figure 1 could be saved and graphed in some other package, although only linear splines are implemented.

The documentation is clear and fairly comprehensive with very few typographical errors (errata sheets provide corrections). There are three manuals: a User's Guide, a Command Summary (both from version 1.8), and a Version 2 Update. The Update includes a master index to all three volumes. The pages for the main manual entries are boldfaced, which is very helpful.

The package is clearly appropriate for routine data analysis, though the lack of high-quality graphics is a drawback. The package does contain some very unusual features, especially the additive and mixed multipicative/additive models. A recently added feature allows the user to adjust `AFMIT`'s Poisson regression results for either over- or under-dispersion, but other than this there is no discussion in the documentation of underlying assumptions and how they might be checked. The package is thus unsuitable for teaching.

6.4 EPILOG PLUS 3

EPILOG is a command–driven program that can be used either interactively or in a batch environment. While some general purpose statistical procedures are included, EPILOG is primarily aimed at analysis of clinical trials and epidemiology and includes many specialized statistics including tests for space–time clustering of events and for clustering of disease within households. Data manipulation capabilities are extensive and are well integrated.

Survival analysis routines, in addition to KM product–limit estimates of survival rates and several tests for comparing groups, are limited to Cox regression, Buckley–James censored regression, and exponential function "cure model" that is limited to one covariable. Both the Cox procedure and the censored regression procedure will automatically form dummy variables and interactions and can be tested for joint significance.

The Cox model can include virtually any type of time–varying covariable and can be stratified. Several plots are available in low resolution on the screen but in high resolution on a graphics printer. However, results cannot be saved or used in any way and the package cannot be extended by the user. Not only is there no general overview of the underlying logic of the program, but the documentation is very brief for each procedure. Several complete examples are presented, but assumptions are not mentioned at all. Up to 50 covariables are allowed in any model.

Because of the extensive examples in the manual, and the built–in editor, the package is fairly easy to learn and to use. However, you must already have a pretty good idea of what you want to do and how you want to do it. EPILOG is certainly suitable for researchers though its limitations and brief manual impede its use for teaching. The low resolution screen plots are also a problem unless the user has a large, high quality, screen. Statistical procedures are not integrated at all which is a further problem for teaching. Further, the on–line help is very skimpy.

6.5 LIMDEP 7.0

LIMDEP is command–driven general–purpose econometrics program that can be used either interactively or batch. The package contains many standard survival analysis techniques, including KM, Cox regression and six parametric models (Weibull, exponential, normal, logistic, Γ, and Gompertz). LIMDEP implements time–dependent covariables in CPH as well as parametric models, and time–varying covariables in parametric models. LIMDEP also includes many specialized techniques included by no other package:

1. The semiparametric model of Han & Hausman (1988) that is oriented toward discrete data with, or without, unobserved heterogeneity. With covariables this is an alternative to Cox's model; without covariables, it is an alternative to KM estimates.

2. Models that account for unobserved heterogeneity across individuals. This is limited to Γ heterogeneity in a Weibull model, and to the Han/Hausman model mentioned above. Unobserved heterogeneity is potentially a very serious problem, much more serious than in linear regression. In regression, omitted variables are not a problem, even when they affect the dependent variable, unless they are correlated with the included covariables. In survival analysis, however, omitted covariables can influence either the parameter coefficients or their standard errors, even if the omitted covariables are uncorrelated with the included covariables (Follman & Goldberg 1988, Lancaster 1990, Yamaguchi 1991). No other package contains any way to deal with this problem directly [2], although LIMDEP's solutions may not be ideal, and theoretically there are a number of other ways to approach the issue.

[2] S-PLUS and Stata can easily compute a covariance matrix estimate that is robust to model misspecification.

3. Truncated data arises when the limit point of 0 is not actually appropriate for all the survival data to be analyzed. For example, when estimating rates of promotion in an organization, one may find that those who left the organization prior to a certain time, say 1986, are not included in the data base, though they may have been promoted and were certainly in competition with those in the data base who were promoted prior to 1986 (Kalbfleisch & Prentice 1988). This is analogous to the truncated ordinary least squares model.

4. Split–population models occur when the standard assumption that all censored cases would eventually fail is false, e.g., the subject is cured. This situation can arise in many social and economic situations; the original article estimated recidivism (Schmidt & Witte 1989), but many other situations quickly come to mind, including promotion, involuntary termination, receipt of a heart transplant, etc.

Prior to version 6, a problem with LIMDEP was the idiosyncratic nature of the command language and the paucity of the manual. The manual has been extensively re–written and lengthened with many more examples. However, tests of assumptions underlying various models are still not discussed. The user interface and help functions have been improved making it much easier to return to the program after an absence of a couple of weeks and easily pick it up again. Also, the graphics have been greatly improved.

The package includes a reasonable data manipulation facility as well as a standard calculator, a matrix language and a general function minimization routine so that users can extend the package without a great deal of trouble. This is further helped by the sizable number of results from each estimation that are retrievable and save-able by the user, and can easily be fed into further program or user–written calculations. Also, tests of linear restrictions on the parameters (joint tests of significance or "partial F–tests") can be set up by the user.

The program is useful for all three purposes discussed above in its latest version. This is particularly true if users have available the author's textbook (Greene 1997) though this is certainly not necessary. The package offers the user a choice of computing algorithms; this is especially useful regarding the parametric models where LIMDEP's default algorithm is different from that offered by the other packages; however, LIMDEP does allow you to switch to the same algorithm offered elsewhere and when you do you obtain the same results as in other packages. The use of different algorithms in, for example, a Weibull model can mean very different results.

6.6 SAS Version 6.10

SAS no longer runs under DOS, but runs under Windows, OS/2, and NT. Parametric survival models with or without covariables are fitted with the `LIFEREG` procedure. The following distributions may be fitted: extreme value, normal, logistic, exponential, Weibull, log–normal, log–logistic, and Γ. For the exponential distribution, an automatic score test of exponentiality vs. Weibull alternatives is given. Left, right, and interval censoring are supported. The procedure is very flexible, having many options except for LR tests and stepwise variable selection. Regression effects for parametric survival distributions are modeled using the accelerated failure time formulation. For exponential and Weibull distributions this is equivalent to the proportional hazards formulation. Most users prefer the proportional hazards form; they will have to go to extra effort to reparameterize the model to obtain, for example, hazard ratios and traditional Weibull parameter estimates. Standard errors and confidence limits for these reparameterizations are very difficult to obtain. Dummy variables are generated automatically for class variables, although many statisticians would prefer that the reference cell be the lowest value instead of the highest (that way a 0–1 class variable will yield the same coefficient as a 0–1 non–class variable). This can be gotten around with the `ORDER` option. Multiple degree of freedom tests are available only for class variables. As a special case, `LIFEREG` will fit binary logistic regression models. Extra input observations with missing censoring indicators may be used to obtain predictions at desired covariable values and follow–up times. An option on the `OUTPUT` statement is used to store predicted survival.

These SAS procedures are extremely flexible, primarily because of the number of options supported as well as the facility of sending analysis results to other procedures. The SAS programming language allows all data transformations. Class and stratification variables may be of any number and type, including character.

For CPH SAS has `PHREG`. It is an outstanding procedure, particularly in regards to how difficult analytic capabilities are implemented. First, `PHREG` has 4 options for handling tied failure times (Breslow, Efron, exact, and discrete logistic). The exact permutation method is efficient for even large datasets with heavy ties. Second, `PHREG` has a nice language for specifying time–dependent covariables.

With time dependent covariables, `PHREG` runs as fast as any CPH program would be expected to do. `PHREG` prints LR, score, and Wald tests and has a TEST statement for pooled Wald tests. It has several variable selection options and will output deviance, martingale, score, Schoenfeld, and scaled Schoenfeld (Grambsch & Therneau 1994) residuals. Predicted survival estimates can be based on the Breslow (cumulative hazard) or Kalbfleisch–Prentice (product–limit) methods. SAS and S-PLUS are the only systems which can compute standard errors for CPH survival estimates.

Some shortcomings of `PHREG` are (1) it does not allow survival predictions at a single point in time or for a given stratum; predictions for all times and all strata are output for each covariable combination, making a dataset that is often too large to fit on the disk. If the time or censoring variable is missing, predictions are missing, thus circumventing the method for getting predictions that is available in all other SAS procedures. Strata by covariable interactions are very problematic here. (2) It does not always print the residual χ^2 statistic for the joint effect of all variables being considered. (3) It does not provide an automatic test of proportional hazards for each predictor as was done with the version 5 SAS procedure PHGLM (Harrell 1986). (4) It mislabels hazard ratio as "risk ratio". (5) Unlike `LIFEREG`, it does not have a `CLASS` statement for generation of dummy variables. (6) `PHREG` will not provide stratified survival estimates if there are no covariables. (7) More emphasis in the documentation is needed on the `MULTIPASS` option for analyzing large datasets with time dependent covariables. The default often causes 'disk full' error messages. (8) The Rao score test is referred to in a historically inaccurate way as the "Lagrange multiplier test".

These non–interactive command–oriented procedures are a standard part of the SAS system. There is no practical limit to data size as SAS stores excess data on disk. Documentation is comprehensive but not always easy to understand and usually has little about testing model assumptions. Full examples are given with annotated output. Computational formulas are given in detail.

SAS is very good for all three purposes listed in section 2, although it takes longer to learn than interactive software and students must go to extra trouble to translate fitted parametric regression models to proportional hazards form. There is a rich variety of data transformations available for parameterizing nonlinear covariable effects. Also, several SAS procedures may be linked to compute log–log KM estimates by quantile groups of continuous variables, for example. These capabilities help form the basis for a complete data analysis strategy.

SAS Institute has withdrawn support for the user–contributed supplemental library. Also, SAS only supports its "Toolkit" product for a couple of platforms for allowing users to write their own procedures. We expect these facts to have ramifications on SAS' ability to make new survival analysis methods available to users in this rapidly evolving field. In addition, the above shortcomings of the `PHREG` procedure demonstrate that SAS does not ask for user input regarding options when new procedures are being designed. Several simple changes in the software would have made it much more flexible.

6.7 SPIDA, Version 6

SPIDA is a command–oriented interactive system that is closest in spirit to the S language. Like S, SPIDA has excellent inter–function communication ability for re–using and re–examining results. The manual has a very nice literature review and is full of excellent examples including some on the use of Cox models for uncensored data. The programming language is flexible and powerful but is annoying in its frequent use of the '$' symbol as a necessary prefix for many variable names. The frequent use of variable numbers rather than names is also a drawback. SPIDA has nice graphics similar to EGRET's. There is a limited facility for automatically handling missing data.

There is one function, `creg`, for parametric modeling of censored responses based on the normal distribution.

The SPIDA functions for unstratified and stratified CPH are separate (`cox` and `scox`). Like other packages, hazard ratios per 1–unit changes in x are printed even when 1–unit is a meaningless increment. The CPH functions implement best–subset and four other variable selection algorithms (the `bcox` and `bscox` functions), collinearity statistics, and influence statistics. The functions provide automatic tests of PH of each predictor as did the SAS Version 5 PHGLM procedure. Displays of model fit are limited. For example, there is no easy way to plot age vs. predicted 2–year survival.

Step–function time–dependent covariables are handled with the `tcox` function. The `tcox` function handles left–truncated data. The manual is truly outstanding in its many descriptions and examples of time dependent covariable and left truncated data analyses.

SPIDA will meet all three needs of Section 2 except for a weakness in the ability to estimate covariable transformations. Its manual alone is a superb teaching tool.

6.8 S-PLUS , Version 3.3

S-PLUS is the enhanced, commercially supported version of the New S language from AT&T (Becker et al. 1988) that runs on Microsoft Windows operating systems. S-PLUS is a highly interactive command–oriented function–based system with the best statistical graphics available as well as hundreds of functions which operate on scalars, vectors, matrices, and more complex S–objects. All data being analyzed are held in memory. Although there is no built–in limitation to the size of dataset that can be analyzed, S is generally a "memory hog" and will require several megabytes of memory to efficiently process datasets with only several hundred subjects.

S-PLUS has the largest number of survival analysis tools of any system, the vast majority of them written by Dr. Terry Therneau of Mayo Clinic. As of version 3.3 all of Therneau's functions are a standard part of S-PLUS .

The `survfit` function computes KM estimates, standard errors (based on survival, log(survival), or log–log survival), and the numbers of events and of subjects at risk. It can also compute the Altschuler–Nelson–Fleming–Harrington–Aalen survival estimator and the Tsiatis standard error estimator. The `plot` method for `survfit` draws stratified survival estimates as multiple curves. The `survreg` function fits logged and unlogged Weibull, logistic, and Gaussian models using the accelerated failure time formulation.

The `coxph` function fits the fixed–covariable CPH model with stratification, and it uses a "start time, stop time" counting process notation to fit time–varying covariables and to fit time–dependent covariables that are approximated by step functions. Time–dependent strata, left truncation, and multiple events per subject are allowed, and subjects may leave the risk set for intervals of time. The manual has good discussion and examples of many of these modeling enhancements. `coxph` returns an S–object with the following components, among others: a frequency table of censoring and missing data by strata, the regression coefficients and covariance matrix, log–likelihood at initial estimates and for the final model, and score test for the model [3]. `coxph` implements three methods of handling tied failure times, and uses a more reasonable default (Efron's method) than the other systems. One of seven types of "residuals" are easily computed with the `residuals` function: martingale, deviance, score, Schoenfeld, scaled (weighted) Schoenfeld, dfbeta, and dfbetas. These residuals can easily be passed through one of S-PLUS 's smoothers and plotted to examine influence of omitted predictors, transformation of predictors, predictor $\times$ time interaction (proportional hazards assumption), and influential observations. The smoothed scaled Schoenfeld residuals (Grambsch & Therneau 1994) are extremely valuable, as such plots estimate the function $\beta(t)$ where β is the regression coefficient that one hopes is not a function of t (i.e., that PH holds). These residuals can easily be used to compute tests of PH, both for all covariables simultaneously and for each predictor separately using a number of possible transformations of the time axis. This is done with the `cox.zph` function.

After running `coxph`, the returned object can be passed to `survfit` to obtain estimated survival probabilities at a given vector of covariable settings. `survfit` will also compute standard errors of predicted Cox survival probabilities. Then the `survfit` output can be passed to `plot` as with Kaplan–Meier estimates.

Also part of S-PLUS are functions for automatically adjusting CPH covariance matrices for model misspecification (e.g., omitted covariables) and intra–cluster correlation, many functions for obtaining adjusted survival curves and survival curves for reference populations, and `survdiff`, which does k–sample survival comparisons using a general family of rank tests as did Fleming, Augustine, Elcombe & Offord (1986).

A major strength of S-PLUS is that all of its regression modeling functions use a standard formula language. Thus unlike SAS, every function not only handles `CLASS` variables but also allows for interactions (of any order) between any types of variables, and it allows for independent variables to be matrices. For example, `age,age`^{2} are always connected (using the formula term `poly(age,2)` if using orthogonal polynomials) so that it's easier to get 2 d.f. tests and hazard ratios for increasing `age`. The only difference between formulas specified to CPH and those specified for parametric models is the absence of `strata` terms for the latter.

[3] Easy access to all components of fit objects makes S-PLUS idea for simulation or bootstrapping. Stata and SPIDA implemented some of this feature.

Therneau has also written S–Plus functions for tree–based (Breiman, Friedman, Olshen & Stone 1984) survival modeling using exponential distributions. These are not part of S-PLUS 3.3.

Survival analysis in S-PLUS is excellent for routine data analysis and teaching, and it is far and away the best environment for model checking and comprehensive survival data analysis. S-PLUS is an ideal environment for the research statistician or the analyst who uses it every day and who wants an extendible system.

The S language is easily extendible. Users can write new functions in S's own high–level language or easily call Fortran or C functions from S functions. Interfacing with other routines is well documented in the New S text (Becker et al. 1988). Thus, existing libraries such as IMSL, NAG, and LINPACK are accessible from S. Many excellent routines are in the public domain in `statlib`. For these reasons, we expect that S-PLUS will continue to make state–of–the–art survival analysis techniques available.

Because S is an object–oriented language with object attributes easily passed between functions, it is possible to make any model–fitting function part of a comprehensive system of tools for creating general design matrices, automatically computing all pertinent tests of pooled effects in the model, plotting model predictions, and many other tasks. More will be said about this in Section 9.

6.9 SPSS, Version 6.1.3

SPSS is primarily a menu-driven package, though one can always use the command language instead, and, for certain things, you must use the command language. This seems a realistic compromise for a professional package that wants to offer a graphical interface. It is either impossible, or places a tremendous burden on the user, to force everything through the menus. While the command language is not easy to learn or use, the user is helped by the practice of building command syntax based on menu choices; in this way, little needs to be done without the menu system and the user is slowly taught the language while producing usable results. A reasonable set of diagnostic graphs and statistics are available including three kinds of residuals and dfbetas (influence statistics).

It is easy to replicate an analysis and it is possible to save much of the output for further processing. There is a matrix language allowing one to extend the package, but it is not well integrated with the built-in routines; you can feed some results into the matrix language, but you are, apparently, expected to build new routines with the matrix language rather than use it to extend the built-in routines.

SPSS has the Cox model, with time–dependent, and time–varying, covariables, and with stratification. No parametric models are offered. The manual includes a discussion of the proportional hazards assumption and methods for checking the assumption both graphically and via the use of time–dependent variables. There is no discussion of the log–linear assumption.

The package is useful for all three purposes discussed above, especially if the students are beginners. The manual is very clear regarding what is available and how to use it. The saved output, e.g. residuals, can easily be passed to other routines or be used to define new graphs. Figure 1, however, could not be easily drawn. Further, anything that is not a straight–forward function of what can be saved would be very hard to calculate.

6.10 Stata, Version 5.0

Stata does CPH as well as Weibull and exponential models and the more general Bailey–Makeham parametric survival model (Bailey, Homer & Summe 1977). Stata is an interactive command–oriented package with good online help, some menu capabilities, batch capability, and excellent easy–to–use graphics. Canned Stata procedures are probably faster than those of any other package. The size of the dataset that can be handled is directly proportional to the size of available memory. By–processing is automatic. Documentation is generally excellent, with good examples. Strong features of all Stata regression models are the abilities to replay the last model fit with a single command and to enter any number of test commands. The test commands by default apply to the last regression model fitted and can specify tests on pooled effects or general contrasts. For models fitted with maximum likelihood, the tests are Wald tests. The use of a generic predict command that applies to a vast array of models is also a nice feature.

Version 5 has completely new routines for survival analysis, so that, although new models were not added, the old models have been significantly enhanced. Additionally, extensive specialized data summary and data management routines have been added that are specifically oriented to survival data. The counting process orientation now underlies the CPH model, allowing: multiple exits and re-entries, gaps, recurring events and delayed entry to the risk pool. Additionally, robust estimates of variance, with or without clustering, has been added to all estimation commands. Diagnostic procedures have also been beefed up with the addition of both Martingale and score residuals. Stratification, on the basis of up to 5 variables, is now allowed for all models.

The program is directly designed for step–function time–varying covariables. However, there is a way to use Stata's data manipulations commands to restructure the data so that time–dependent covariables can be fit. Though not discussed in the manual, there is a FAQ (frequently asked questions) at their World Wide Web site (`http://www.stata.com`) on how to do this.

Stata has a great deal of flexibility with respect to making predictions; e.g., one can predict for any combination of the covariates on either the estimation sample or on a new sample.

Stepwise model fitting is allowed for any model. Breslow's likelihood is used for tied survival times.

Included in the new routines are several descriptive routines that are specifically oriented to survival–type data and others that are specifically aimed at programmers who want to extend Stata's survival analysis capabilities. For example, in one convenient command, one can now learn easily not only how many subjects and how many failures are present, but also how many subjects have gaps and how long these gaps are, what the distribution of time at risk is for the total population (or subsets of it), and information about delayed entry subjects. Another new command provides summary statistics including the time at risk, the incidence rate, and some quantiles of survival time. A third example is a command that reports which covariates vary within subject over time; the pattern of missing values for such variables is also presented (and there are other commands for imputing replacement values). Note that Stata also has special descriptive statistics commands for other specialized forms of data, such as longitudinal (or, panel) data.

Another addition is a special command to generate new variables that reflect entire histories for each subject. These histories might include the number of failures, the number of gaps, the lengths of gaps, etc.

Other new commands interface to other types of analysis within Stata. For example, one command provides incidence rate comparisons.

All statistical commands save their results into specialized macros that can be used in further programming, can be passed to other Stata procedures, or can be used to change the format of the output. Stata has good transformation functions (including several ways to nearly automatically generate dummy variables and interaction terms) and a good language for implementing user extensions. Unlike most other programmable packages, there is a standard way for Stata users, who are willing to pay, to receive information about, and copies of, new Stata routines written by other users, or by the vendor: the Stata Technical Bulletin (STB). New commands have been added in this way. An example is the cubic spline routine in the STB (Goldstein 1992), which can be used to form spline functions that allow checking of the functional form for any single–equation estimation command in Stata, including Cox (`stcox`), Weibull (`stweib`) and exponential (`stereg`) survival models. The command automatically forms the needed variables using 3–7 knots and then tells the user how to test for possible problems, both graphically and statistically. Another source of extensions is the Stata listserver, which also sometimes includes vendor–written extensions. A recent example, requested by one user, is Harrell's *c*–statistic, used in quantifying the predictive ability of a fitted survival model (Harrell, Lee & Mark 1996). Like `Statlib` and S–news for S-PLUS , the STB and the listserver encourages the user community to quickly extend the package as new methods develop.

Stata is good for routine data analysis and is an excellent teaching tool. For comprehensive data analysis and model checking, Stata is good in some areas. For example, there is some discussion of, and some tools for testing, the proportional hazards assumption, though more could be done. However, there is no discussion of checking the assumption of log–linearity.

6.11 Statistica 5.0

Statistica is a menu–driven package, with an accompanying command language for both batch runs and for extensions to the package. The command language is neither easy to learn nor easy to use. The package is modular; the data management and graphics capabilities are well integrated with the statistical procedures, but the various statistical procedures are not well integrated. The graphs provided are of very high quality; however, we found it impossible to obtain or design many graphics that are considered standard in survival analysis (e.g., log(-log(survival)) against time); graphs such as Figure 1 are not available either. No residuals or other diagnostic statistics (e.g., delta–beta's, predictions, etc.) are available for the Cox model.

It is easy to replicate an analysis and, with a little work, it is possible to save all output from an analysis for further processing. The biggest problem here was that the amount of output available was strictly limited, putting a strain on the user who must form desirable quantities (such as various kinds of residuals) without any help from the software or the manual.

Statistica offers the Cox model (with time–dependent, but not time–varying, covariables), and three parametric models (exponential, normal and log–normal). Stratification is available with CPH.

The manual does supply a brief discussion of both assumptions underlying the Cox model: the proportional hazards and log–linear assumptions. A procedure for testing the PH assumption is provided via the use of a time–dependent variable. However, in general the manual is too brief and too simplistic to be useful in teaching; an accomplished expert in survival analysis could use the underlying tools to produce a reasonable analysis, but why would anyone want to program in various statistics and graphs that are already widely available in other packages?

6.12 SURVIVAL, Version 3.0

This is a command–driven package that can be used either interactively or batch. This package only performs various kinds of survival analysis. However, virtually all of the output can be saved and used in other software packages (including other packages from Salford Systems). The package cannot be extended; data manipulation capabilities are fair and fully integrated with the analyses.

Both Cox proportional hazards and parametric models (exponential, Weibull, log–normal and log–logistic) are provided. The Weibull and exponential models can be estimated in either proportional hazards or accelerated failure time form. This is the only package that can handle interval–censored data in parametric models; this is accomplished via Turnbull's extension to KM.

Stratification and time–varying covariables are allowed in CPH, but not parametric models; time–dependent covariables are not allowed for any model.

The built–in graphics are wide–ranging, but not of high quality; the data underlying each graph can be saved and exported to other packages. As with BMDP, a nice graph of censoring patterns is available.

Predictions are available and can be both graphed and saved. The model coefficients and covariance matrix can also be saved; all of this can easily be exported to other packages. There is also an option for saving Cox–Snell residuals, though there is no mention of other types of residuals.

The manual is very good and clear with many examples; discussion of assumptions is limited to the PH assumption and graphical procedures for examining whether this was violated. The manual includes a very good discussion of types of censoring; further, this is the only manual that discusses the assumption that the censoring not be informative; a discussion of what informative censoring is and how it might arise is included.

This package is valuable for all three uses discussed above, though more emphasis in the manual on the meaning of the models and how to examine the underlying assumptions would be valuable. For teaching purposes, a particularly valuable part of the manual discusses why the results from this package might differ from published results (and from other packages).

6.13 TRUE EPISTAT, Version 5.3

This menu–driven package is aimed at health and medical researchers; within the limits of what the vendor thinks these researchers are interested in, this is a general–purpose package. One entire section is for survival analysis, including Cox proportional hazards models, but not including parametric models. The package includes a macro capability but no ability to extend the package's capabilities.

TRUE EPISTAT is very easy to learn, but, because of the heavy use of jargon, it is not always easy to return to if you have been away from it for awhile. While there are numerous examples in the manual, they are generally not complete and so are less help than they could be.

There is on–line help; also, the main menu contains, next to each menu choice, the pages in the manual that describe that choice. If you are a regular user, then the use of macros can make TRUE EPISTAT very easy to use. TRUE EPISTAT automatically forms dummy variables from all categorical variables and automatically provides a joint test of significance of the set of dummy variables.

The manual discusses some of the underlying assumptions, including some not generally mentioned by other packages, but does not discuss others (e.g., no mention is made of the log–linear assumption). The documentation contains glaring errors, e.g. "Life Table Analysis is not appropriate if there are large numbers of subjects still alive (no Event) at the end of the study ...".

Predicted probabilities can be saved, and easily graphed in various ways. Residuals cannot be saved, though it is easy to calculate certain types of residuals using the package's transformation capabilities, if you have saved the predicted probabilities, and if you know what you are doing—the manual provides no guidance. TRUE EPISTAT can deal with internal time varying covariables of the following types: (a) initially absent but later appears, such as metastases; (b) initially present but then disappears due to, for example, surgery; (c) present at one level and then changes level. Only one change is allowed for each time–varying covariable. External time varying covariables (Kalbfleisch & Prentice 1980) are not allowed, although the reason for this is not clear.

This program is certainly useful to researchers, though it is unclear how useful it is for either teaching isolated components of the theory underlying a given tool or for teaching how a given tool is part of a complete analysis strategy. The manual does provide citations and formulae, but the program is not really flexible enough to accommodate some teachers' desires to break a tool down to its component parts.

Because of the statistical misunderstandings in the manual, we do not recommend this package for students or non–expert practitioners.

7 SUMMARY TABLE

The table below summarizes our comparisons. The following abbreviations are used:

- under "Nonparametric":
  - "K–M" is the Kaplan–Meier product limit estimator
  - "N–A" is the Nelson–Aalen estimator
- under "Semi–Parametric":
  - "CPH" is Cox's proportional hazards model
  - "t–d" means that one can use time–dependent variables in CPH
  - "t–v" means that one can use time–varying (or, step function) variables in CPH
  - "H–H" means the Han–Hausman estimator (Han & Hausman 1988)
- under "Types of Residuals" (discussion of the various residuals for CPH models can be found in Grambsch & Therneau (1994);Therneau, Grambsch & Fleming (1990); Collett (1994)); note that the Martingale and Schoenfeld residuals appear to be the most important types as they are particularly useful for checking major regression assumptions; the Score residuals are useful in checking for influential observations, as well as for computing sandwich covariance matrix estimates:
  - "C–S" is Cox–Snell

  - “Std” means standardized residuals
  - “M” is martingale residuals; these are useful for assessing the adequacy of an hypothesized transformation of a predictor variable
  - “D” is deviance residuals
  - “Score” is score residuals
  - “Sch” is Schoenfeld residuals which are useful for testing the PH assumption

- under “Doc[umentation] for checking assumptions”, the possible entries are:
  - “Yes”, meaning the testing of both the proportional hazards (PH) assumption and the log–linear assumption are discussed in the manual
  - “PH” meaning that only the testing of the PH assumption is discussed
  - “no” meaning there is no discussion of testing assumptions in the manual; the user may still be able to test assumptions, but the manual provides no help or guidance

- under “Left–Trunc[ation]–Delayed Entry”:
  - “DE” means the package can handle observations whose entry into the risk set is delayed
  - “LT” means the package can handle left–truncated data
  - “neither” means that the package can’t deal with either delayed entry or left–truncation

- while there are no abbreviations under “Type of Package”, it is worthwhile to note that packages that are extendable, and have active user–communities making extensions, may well have additional techniques; two of the included packages have such user–communities:
  - S-PLUS ; not only was the original routine written by a user (Terry Therneau), but the second author of this paper has written a coordinated set of functions for in–depth survival modeling which includes:
    * relaxing and testing linearity assumptions using cubic splines
    * automatic pooled Wald tests
    * bootstrap model validation
    * presentation graphics for describing models

    See Harrell et al. (1996) and Harrell (1996)
  - Stata; there were many additions to the prior version, including Bailey–Makeham models, which will also work with the current version; the first author has added cubic splines to Stata’s toolbox.

Package	Nonparametric	Semi–Parametric	Parametric	Types of Residuals	Doc. for Checking Assumptions	Counting Process	Left Trunc./ Delayed Entry	Type of Package
BMDP	K–M	CPH, t–d, t–v	several	C–S, Std	Yes	No	DE	complete, general
Egret	K–M	CPH, t–d, t–v	several	none	PH	No	DE	special purpose
Epicure	K–M, N–A	CPH, t–d, t–v	several	M, D, Score, Sch	PH	Yes	DE	special purpose
Epilog Plus	K–M	CPH, t–d, t–v	none	none	PH	No	neither	complete, biostat
Limdep	K–M	CPH, t–d, t–v, H–H	several	none	PH	No	LT	complete, econometric
SAS	K–M	CPH, t–d, t–v	several	M, D, Score, Sch	PH	Yes	DE	complete, general
Spida	K–M	CPH, t–d, t–v	none	none	PH	No	DE	complete, general
S-PLUS	K–M, N–A	CPH, t–v	several	M, D, Score, Sch	Yes	Yes	DE	complete, general
SPSS	K–M	CPH, t–d, t–v	none	C–S, M, Sch	PH	No	neither	complete, general
Stata	K–M	CPH, t–d, t–v	Exp, Weib	M, Score	PH	Yes	DE	complete, general
Statistica	K–M	CPH, t–d	several	none	PH	No	neither	complete, general
Survival	K–M	CPH, t–v	several	none	PH	No	neither	special purpose
True Epistat	K–M	CPH, t–v	none	none	no	No	neither	complete, biostat

8 ADVANCED SURVIVAL ANALYSIS CAPABILITIES

Very recently, packages have begun to implement more advanced survival analysis techniques as described below.

1. Estimation of $S(t)$ in the presence of time–dependent covariables can be a difficult task. EGRET, EPICURE, SPIDA, and S-PLUS do this.
2. Confidence limits for estimated survival probabilities for individual subjects ($\hat{S}(t)$) require the use of somewhat messy variance formulas. SAS and S-PLUS provide these confidence intervals, S-PLUS doing so even when (step–function) time–dependent covariables are present.
3. Smooth nonparametric estimates of the underlying hazard function can be of great assistance in understanding the mechanism of failure. EGRET and EPICURE implement kernel hazard estimates, even when there are time–dependent covariables.
4. The Andersen & Gill (1982) "counting process" model is a powerful generalization of the CPH model which is very flexible. This model allows for (1) multiple events per subject, (2) disjoint observation intervals, left trunction, and other means of controlling the constitution of the risk set, and (3) time–dependent stratification. SAS, S-PLUS , and Stata have general implementations of the Andersen–Gill model.
5. Another advanced feature is implementation of frailty models which explicitly allow unspecified heterogeneity across individuals (which may arise because of unmeasured covariables) or correlations between multiple events within individual. LIMDEP is the only package with such features.
6. EPICURE implements likelihood–based confidence intervals and has a variance adjustment for use with case–cohort designs.

9 DISCUSSION: THE IMPORTANCE OF SOFTWARE INTEGRATION

BMDP, EGRET, EPICURE and other products perform certain dedicated survival analysis tasks very well. However, these products have two very different drawbacks that may prevent them from accomplishing all steps of a thorough data analysis. The BMDP survival analysis programs are part of a general system that allows user programming and general transformations of predictor variables. However, BMDP as a whole does not provide for good inter–communication between programs via standard format datasets. The BMDP program for CPH for example will not allow user–specified predictions to be output to a file for further analysis and plotting.

EGRET, on the other hand, is a self–contained package. It has some of the same data communication shortcomings as BMDP, but a bigger problem is that it does not provide comprehensive descriptive statistics nor does it have a programming language or algebraic processor to allow critical examination of regression assumptions. Cubic spline expansions of `X` are not on EGRET's menu.

EPICURE, though not a menu–driven program, has many similarities to EGRET on this score. So does SURVIVAL.

Figure 1, described in Section 4, demonstrates the value of being able to request predictions for computed values (e.g., quartiles of a covariable) and passing datasets between procedures, as well as the value of SAS's macro language. The capabilities of the S-PLUS statistical modeling language (Chambers & Hastie 1992) allows even more general modeling functions without needing a macro language. FH has written a library of S-PLUS functions for UNIX or Windows called `Design` (Harrell 1996) that work with the modeling language functions and use Therneau's survival functions built in to S-PLUS . All pertinent tests such as linearity and factor tests (main effects pooled with interaction effects containing that factor) are performed automatically. For example, suppose one wished to fit a Cox model with the following candidate predictors: age, modeled with a restricted cubic spline function (using default knot placement); treatment, a 3–group categorical variable; and white blood count, known to require the transformation $\log(x + 1000)$, race, and sex. The latter two variables are used as stratification factors so as to not assume proportional hazards. An S-PLUS program to fit the model, allowing 2–way interactions between treatment and sex and between age and race, with a general smooth age effect follows. Note that not all strata (sex $\times$ race cross–classifications) need interact with treatment and age.

```
fit ← cph(Surv(dtime, death) ~ log(wbc+1000) +
          treat*strat(sex) + rcs(age)*strat(race), surv=T )
 specs(fit)              #describe design, e.g. knot placement
 anova(fit)              #automatic contrasts
 summary(fit,treat="C")  #hazard ratios, "C"=reference cell
 plot(fit,age=NA,wbc=NA,treat="C") #NA:default range
 survplot(fit,treat=NA, n.risk=T)
 survest(fit, expand.grid(age=c(40,60),wbc=9000,treat=c("A","B"),
        sex="male", race="black"), times=c(2,5), conf.int=.99)
 fastbw(fit)             #fast step-down variable selection
 xb ← Function(fit)      #compose S function to compute X*beta
                         #analytically as needed
 xb(wbc=30000, treat='A', sex='male', age=50, race='black')
                         #quickly obtain X*beta for set of characteristics
 sascode(xb)             #translate formula for predictions to SAS
 nomogram(fit)           #draw a nomogram for getting predictions manually
 latex(fit)              #typeset model fit
```

The `cph` function fits the model and returns information in a "fit object". The object can optionally include underlying survival estimates and their standard errors. The `anova` function prints among other things Wald statistics for the following effects: 3 factor effects (main + interaction effects), 2 interaction effects, test of linearity for age, all interactions, linearity of interaction between age and race, and combined linearity of age and interactions with age. `summary` prints hazard and log hazard ratios and confidence intervals, taking into account interactions and nonlinearity — by default inter–quartile hazard ratios are computed for continuous variables. Here treatment C is considered the reference group instead of the default, treatment A. The `plot` call produces a 3–dimensional perspective plot of the joint effect of age and wbc (original scale, not log(wbc+1000) scale) on log hazard, adjusted to treatment=C. The range for plotting is by default the 10th smallest to 10th largest value of a covariable if $n > 200$. `survplot` plots survival curves, with time on the x–axis and the number of subjects at risk for each time interval placed above the x–axis. The `survest` function estimates 2– and 5–year survival probabilities and 0.99 confidence limits for a variety of covariable and strata combinations. `fastbw` does a Lawless & Singhal (1978) backward step–down analysis. The `latex(fit)` command stores and displays a typeset document showing the fitted model in mathematical form using LaTeX (Lamport 1994). Since Therneau has also implemented a set of functions for parametric survival modeling in S-PLUS , replacing `cph` with `psm` will fit an exponential, Weibull, log–logistic, or log–normal survival model (except for stratification) and all of the functions called after `cph` would remain the same. Likewise, it is easy to fit logistic, Poisson, and other models and still use all of the other functions that process the "fit object". A function in `Design` called `pphsm` will convert a Weibull or exponential model from accelerated failure time form to hazard ratio form. Other functions create S-PLUS functions to analytically compute survival probabilities, life expectancy, quantiles, and hazard functions for individual subjects from any of the parametric models.

To obtain a bootstrap validation of the accuracy of 2–year survival estimates which attempts to replicate the variables selected with step–down selection at each of B=100 bootstrap samples, the parameters `x=T, y=T, time.inc=2` would be added to the `cph` or `psm` invocation, and the command

```
calibrate(fit, bw=T, u=2, B=100)    #u=2 yrs., bw=backward step-down
```

would be run.

See Harrell et al. (1996) for a case study in survival modeling that includes detailed examples of the use of the `Design` library with output.

10 CONCLUSIONS

There is PC survival analysis software that is worth buying. A classic trade–off exists however in choosing which software to purchase: software with the best interactive user interface can also be the least flexible. A potential buyer may want to contrast the extra hours of learning a flexible command–oriented program to the future flexibility offered by the program. The same issues are raised in choosing software to be used in survival analysis courses. Some software can be easily learned by students and will provide insight into isolated components of survival analysis. However, that software will not always teach the student good data analysis schemes that includes critical examination of regression assumptions and fitting necessary non–linear or non–additive effects. Survival analysis software that is part of a general statistical package having a programming language, macro language, and input/output dataset options will be the most useful software in general. In addition, if the package also runs on larger computers, it provides a software bridge allowing larger problems to be quickly moved to the larger computer without reprogramming or learning a new language. S-PLUS , Stata, and SPIDA are some of the newer packages that combine excellent graphics with a good interactive command language and good inter–program communication of results. The very high level object–oriented S language demonstrates what is currently possible in achieving all goals except for instant point–and–click usability.

References

Altman, D. G. & De Stavola, B. L. (1994), "Practical problems in fitting a proportional hazards model to data with updated measurements of the covariates", *Statistics in Medicine* **13**, 301–341.

Andersen, P. K. & Gill, R. D. (1982), "Cox's regression model for counting processes: A large sample study", *Annals of Statistics* **10**, 1100–1120.

Bailey, R. C., Homer, L. D. & Summe, J. P. (1977), "A proposal for the analysis of kidney graft survival", *Transplantation* **24**, 309–315.

Becker, R. A., Chambers, J. M. & Wilks, A. R. (1988), *The New S Language*, Wadsworth and Brooks/Cole, Pacific Grove CA.

Breiman, L., Friedman, J. H., Olshen, R. A. & Stone, C. J. (1984), *Classification and Regression Trees*, Wadsworth and Brooks/Cole, Pacific Grove, CA.

Chambers, J. M. & Hastie, T. J., eds (1992), *Statistical Models in S*, Wadsworth and Brooks/Cole, Pacific Grove CA.

Clayton, D. (1988), "Review of SURVCALC", *Applied Statistics* **37**, 455–456.

Collett, D. (1994), *Modelling Survival Data in Medical Research*, Chapman and Hall, London.

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

Dain, B. J., Freeman, D. H. & Vredenburgh, J. J. (1989), Comparison of different packages' survival test results, *in* "Proceedings of the Statistical Computing Section ASA", pp. 315–318.

Fleming, T. R., Augustine, G. A., Elcombe, S. A. & Offord, K. P. (1986), The SURVDIFF procedure, *in* "SUGI Supplemental Library User's Guide, Version 5 Edition", SAS Institute, Inc., Cary NC, pp. 535–553.

Follman, D. A. & Goldberg, M. S. (1988), "Distinguishing heterogeneity from decreasing hazard rates", *Technometrics* **30**, 389–396.

Goldstein, R. (1991), "Statistical Computing editor's notes", *American Statistician* **45**, 304–305.

Goldstein, R. (1992), "Restricted cubic spline functions", *Stata Technical Bulletin* **STB-10**, 29–32.

Goldstein, R., Anderson, J., Ash, A., Craig, B., Harrington, D. & Pagano, M. (1989), "Survival analysis software on MS/PC–DOS computers", *Journal of Applied Econometrics* **4**, 393–414.

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

Greene, W. H. (1997), *Econometric Analysis*, third edn, Prentice Hall, New York.

Hallahan, C. (1992), "DBMS/COPY and DBMS/COPY Plus (Version 2.0)", *American Statistician* **46**, 49–52.

Han, A. & Hausman, J. (1988), Semiparametric estimation of duration and competing risk models, MIT photocopy.

Harrell, F. E. (1983), Data analysis considerations and survival analysis programs, *in* J. E. Gentle, ed., "Computer Science and Statistics: The Interface", North-Holland, Amsterdam, pp. 93–96.

Harrell, F. E. (1986), The PHGLM Procedure, *in* "SUGI Supplemental Library Users Guide", Version 5 edn, SAS Institute, Inc., pp. 437–466.

Harrell, F. E. (1996), "`Design`: S–Plus functions for biostatistical/epidemiologic modeling, testing, estimation, validation, graphics, prediction, and typesetting by storing enhanced model design attributes in the fit. UNIX and Microsoft Windows versions available from `http://lib.stat.cmu.edu`.".

Harrell, F. E., Lee, K. L. & Mark, D. B. (1996), "Multivariable prognostic models: Issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors", *Statistics in Medicine* **15**, 361–387.

Harrell, F. E., Lee, K. L. & Pollock, B. G. (1988), "Regression models in clinical studies: Determining relationships between predictors and response", *Journal of the National Cancer Institute* **80**, 1198–1202.

Harrell, F. E., Pollock, B. G. & Lee, K. L. (1987), Graphical methods for the analysis of survival data, *in* "Proceedings of the Twelfth Annual SAS Users Group International Conference", SAS Institute, Inc., Cary, NC, pp. 1107–1115.

Hoffman, R. G. (1985), Simulated temporal models for proportional hazards models with time varying covariates, *in* "Proceedings of the Statistical Computing Section ASA", pp. 390–395.

Hudec, M., Hitz, M. & Muellner, W. (1985), Software for survival analysis. A comparative survey (German), *in* "Proceedings of the Statistik-Software Third Conference on the Scientific Use of Statistical Software", Gustav Fischer Verlag, Stuttgart, pp. 203–213.

Huster, W. & Enterline, J. (1985), "SURVPAK: an interactive survival analysis package", *American Statistician* **39**, 311.

Iwane, M., Palensky, J. & Plante, K. (1997), "A user's review of commercial sample size software for design of biomedical studies using survival data", *Controlled Clinical Trials* **18**, 65–83.

Kalbfleisch, J. D. & Prentice, R. L. (1980), *The Statistical Analysis of Failure Time Data*, Wiley, New York.

Kalbfleisch, J. D. & Prentice, R. L. (1988), "Estimation of reliability in field-performance studies, with discussion", *Technometrics* **30**, 365–388.

Kaplan, E. L. & Meier, P. (1958), "Nonparametric estimation from incomplete observations", *Journal of the American Statistical Association* **53**, 457–481.

Lamport, L. (1994), *LaTeX: A Document Preparation System*, second edn, Addison–Wesley, Reading, MA.

Lancaster, T. (1990), *The Econometric Analysis of Transition Data*, Cambridge University Press, Cambridge.

Lawless, J. F. (1982), *Statistical Models and Methods for Lifetime Data*, Wiley, New York.

Lawless, J. F. & Singhal, K. (1978), "Efficient screening of nonnormal regression models", *Biometrics* **34**, 318–327.

MathSoft (1995), *S-Plus User's Manual, Version 3.3*, MathSoft, Inc., Seattle WA.

Schmidt, P. & Witte, A. D. (1989), "Predicting criminal recidivism using 'split population' survival time models", *Journal of Econometrics* **40**, 141–159.

Therneau, T. M., Grambsch, P. M. & Fleming, T. R. (1990), "Martingale-based residuals for survival models", *Biometrika* **77**, 216–218.

Wagner, A. E. & Meeker, W. Q. (1985), A survey of statistical software for life data analysis, *in* "Proceedings of the Statistical Computing Section, ASA", pp. 441–446.

Yamaguchi, K. (1991), *Event History Analysis*, Sage, Newbury Park.

Appendix
Vendors of Software Reviewed

BMDP/PC Version 7
SPSS Inc.
444 N. Michigan Ave.
Chicago, IL 60611
(312) 329–4000

EGRET 1.02.10
Statistics and Epidemiology Research Corp.
1107 NE 45^{th} Street
Suite 520
Seattle, WA 98105
(206) 632–3014 FAX:(206) 547–4140

EPICURE 2.0
HiroSoft International Corp.
1463 E. Republican Ave.

Suite 103
Seattle, WA 98112
(206) 328–5301

EPILOG PLUS 2.10
EPICENTER SOFTWARE
P.O. Box 90073
Pasadena, CA 91109
(818) 304–9487

LIMDEP 7.0
Econometric Software, Inc.
15 Gloria Place
Plainview, NY 11803
(516) 938–5254

SAS Version 6.10
SAS Institute, Inc.
Box 8000
Cary NC 27511
SAS Institute Inc. (1990):SAS/STAT User's Guide, Version 6,
Fourth Edition, Volume 2. Cary NC: SAS Institute, Inc.
SAS Institute Inc. (1994):SAS/STAT Software: Changes and Enhancements, Release 6.10, Cary NC: SAS Institute, Inc.
(919) 677–7000

SPIDA Version 6
The Statistical Laboratory
Macquarie University
NSW 2109 Australia
02–850–8792
fax: 02–850–7285

S-PLUS Version 3.3
MathSoft, Inc., Data Analysis Products Division
1700 Westlake Avenue North, Suite 500
Seattle WA 98109
(206) 283–8802

SPSS Version 6.1.3
SPSS Inc.
444 N. Michigan Ave.
Chicago, IL 60611

(312) 329–4000

Stata Version 4.0
Stata Corporation
702 University Drive East
College Station
Texas 77849
(800) 782–8272

Statistica 5.0
StatSoft
2325 East 13th Street
Tulsa, OK 74104
(918) 583–4149

Survival 3.0
Salford Systems
8880 Rio San Diego Dr., Suite 1045
San Diego, CA 92108
(619) 543-8880

TRUE EPISTAT 5.3
Epistat Services
2011 Cap Rock Circle
Richardson, TX 75080–3417
(972) 680–1376