31
Survival Analysis

Paul D. Allison

Survival analysis is a collection of statistical methods that are used to describe, explain, or predict the occurrence and timing of events. The name *survival analysis* stems from the fact that these methods were originally developed by biostatisticians to analyze the occurrence of deaths. However, these same methods are perfectly appropriate for a vast array of social phenomena including births, marriages, divorces, job terminations, promotions, arrests, migrations, and revolutions. Other names for survival analysis include *event history analysis*, *failure time analysis*, *hazard analysis*, *transition analysis*, and *duration analysis*. Although some methods of survival analysis are purely descriptive (e.g., Kaplan-Meier estimation of survival functions), most applications involve estimation of regression models, which come in a wide variety of forms. These models are typically very similar to linear or logistic regression models, except that the dependent variable is a measure of the timing or rate of event occurrence. A key feature of all methods of survival analysis is the ability to handle *right censoring*, a phenomenon that is almost always present in longitudinal data. Right censoring occurs when some individuals do not experience any events, implying that an event time cannot be measured. Introductory treatments of survival analysis for social scientists can be found in Teachman (1983), Allison (1984, 1995), Tuma and Hannan (1984), Kiefer (1988), Blossfeld and Rohwer (2001), and Box-Steffensmeier and Jones (2004). For a biostatistical point of view, see Collett (2003), Hosmer and Lemeshow (2003), Kleinbaum and Klein (2005), or Klein and Moeschberger (2003). Specific desiderata for applied studies that use survival analysis are presented in Table 31.1 and later explained in detail.

1. Definition of the Event

The first step in any application of survival analysis is to define, operationally, the event that is to be modeled. Ideally, an event is a qualitative change that occurs at some specific, observed point in time. Classic examples include a death, a marriage, or a promotion. In such cases, where there is little ambiguity, there may be no need to explicitly define the event. Other applications may not be so clear cut, however. Some changes (e.g., menopause) take a while to "occur," so it is necessary to make decisions about criteria for determining the timing of the event. It is also possible to define events with respect to quantitative variables, especially if they undergo sharp, sudden changes. For instance, a "stock

Table 31.1 Desiderata for Survival Analysis

Desideratum	*Manuscript Section(s)**
1. The event is defined in a clear and unambiguous way.	I
2. The observation period is specified with careful consideration of origin time and possible late entry.	M
3. Censoring is discussed, with indications of amount, type and reasons for censoring.	M
4. An appropriate choice is made between a discrete versus a continuous time method.	M
5. An appropriate choice is made between a parametric versus a semi-parametric method.	M
6. Choice of covariates is discussed and justified. Possible omitted covariates are considered.	M, D
7. Any time-varying covariates are appropriately defined, and a method for handling them is chosen.	M
8. If there are multiple events per individual, an appropriate method is chosen to handle the possible dependence among those events.	M
9. If there are competing risks, an appropriate method is chosen and appropriate tests are reported.	M
10. Sampling method and sample size are explained and justified.	M
11. The treatment of missing data is addressed.	M, R
12. The name and version of the software package is reported.	M,R
13. Summary statistics of measured variables are presented; information on how to gain access to the data is provided.	R
14. Graphs of the survivor function(s) are presented.	R
15. The proportional hazards (or equivalent) assumption is evaluated.	R
16. For competing models, comparisons are made using statistical tests (for nested models) or information criteria (for non-nested models).	R
17. Coefficients (or hazard ratios) are reported, together with standard errors, confidence intervals and *p*-values.	R
18. Conditional survivor and/or hazard functions may be presented.	R
19. Potential methodological limitations are discussed.	D

* *Note*: I = Introduction, M = Methods, R = Results, D = Discussion

market crash" could be said to occur if a particular market index falls more than 30% during a single week. Clearly, this definition involves some arbitrary choices that must be carefully considered and justified. A person could be said to "fall into poverty" if his income falls below some specified threshold. But this demands a rationale for choosing that threshold.

Another decision that must be made is whether to treat all events the same or to distinguish different types of events. If the event is an arrest, for example, one could either treat all arrests the same or distinguish between arrests for misdemeanors and arrests for felonies. All deaths could be treated alike, or one could distinguish between different kinds of deaths according to reported causes. Of course, such distinctions are only possible if data are available to differentiate the event types. Why do it? Usually, it is done because there are reasons to believe that predictor variables have different effects on different event types. In such cases, the prevailing strategy is to estimate competing risks models (see Desideratum 9). The downside of distinguishing different event types is that fewer events are available to estimate each set of parameters, which might substantially reduce statistical power.

Lastly, when events are repeatable for each individual, one must decide whether to focus on a single (usually the first) event for each individual, or to use a method that incorporates all the repeated events. If the average number of events per individual is small, say, less than two, it is usually better to restrict attention to the first event.

2. Observation Period

Survival analysis requires that each individual be observed over some defined interval of time; if events occurred during that interval, their times are recorded. If events are not repeatable, observation is often terminated at the occurrence of an event. Decisions about the starting and stopping times for the observation period should be reported and justified.

Most methods of survival analysis (e.g., Cox regression) require that the event time be measured with respect to some *origin time*. The choice of origin time is substantively important because it implies that the risk of the event varies as a function of time since that origin. In many cases, the choice of origin is obvious. If the event is a divorce, the natural origin time is the date of the marriage. In other cases, the choice is not so clear cut. If the event is a retirement, do you model age at retirement or time in the labor force?

Ideally, the origin time is the same as the time at which observation begins, and most software programs for survival analysis presume that this is the case. Frequently, however, observation does not begin until some time after the origin time. For example, although we may use date of marriage as the origin time in a study of divorce, couples may not be recruited into the study until years later. This is called *late entry* or *left truncation*. Because individuals are not at risk of an observed event until observation begins, special methods are necessary to take this into account. For more details, see Allison (1995, pp. 161–165)

3. Censoring

Censoring is endemic to survival analysis data, and any report of a survival analysis should discuss the types, causes, and treatment of censoring. By far the most common type of censoring is *right censoring*, which occurs when observation is terminated before an individual experiences an event. For example, in a study of divorce, couples that do not divorce during the observation period are right censored. All survival analysis software is designed to handle right censoring, and it is essential to include the right censored observations in the analysis.

Standard methods for dealing with right censoring presume that such censoring is *non-informative*. Roughly speaking, that means that the fact that an individual is censored at particular point in time does not tell us anything about that individual's risk of the event. That assumption is necessarily satisfied if the censoring time (or potential censoring time) is the same for everyone in the sample. However, the censoring could be informative if it occurs at varying times because individuals drop out of the study, which could lead to biased estimates of the parameters. Unfortunately, there is no test for the non-informative assumption and little that can be done to correct for bias due to violation of this assumption. But the lesson is that survival studies should be designed and executed so as to minimize censoring due to drop outs. In any case, the proportion of censoring cases due to drop outs should be reported.

A slightly less common type of censoring is *interval censoring*, which means that an individual is known to have an event between two points in time, but the exact time is unknown. For example, if a person reports being unmarried at wave 1 of a panel study but married at wave 2, then the marriage time is interval censored. If the censoring times are regularly spaced, interval censoring can often be handled by discrete-time methods (see the next section). However, most survival analysis software cannot handle irregular patterns of interval censoring.

The least common type of censoring is *left censoring*, which happens when an event is known to have occurred before some particular time, but the exact time is unknown. For example, in a study of first marriage, if a person is known only to have married before age 20, that person's marriage age is left censored. Note that the term left censoring is often used with a quite different meaning in the social science literature. In this alternative meaning, left censoring is said to occur when we begin observing an individual at some arbitrary point in time, but we do not know the origin time (i.e., how long it has been since the individual has been at risk of the event).

4. Discrete-Time vs. Continuous-Time Methods

If you know the exact times at which events occur, it is appropriate to use methods that treat time as continuous. If, on the other hand, you know only the month or the year of the event, you might be better off using discrete-time methods. One of the best indications of the need for discrete-time methods is the presence of large numbers of *ties.* A tie is said to occur if two individuals experience an event at the same recorded time. Occasionally, time is truly discrete in the sense that events can only occur at certain discrete points in time. For example, in most universities, faculty can only be promoted at the end of an academic year.

Most survival analysis software is designed for continuous-time data. If you want to go the discrete-time route, you must choose between a logit model and a complementary log-log model. Logit is more appropriate for event times that are truly discrete, while complementary log-log is more appropriate for events that can happen at any time but are only observed to occur in discrete intervals. In practice, the choice is usually not consequential.

Having chosen a model, you must then choose an estimation method. Some Cox regression programs (e.g., SAS, Stata, S-Plus) have options for estimating either of the two models using partial likelihood estimation. But partial likelihood can be very computationally intensive for large samples with lots of ties. The alternative is to do maximum likelihood using conventional binary regression software. The trick is to break up each individual's event history into a set of distinct records, one for each unit of time in which the individual is observed, with a dependent variable coded 1 if an event occurred in that time unit, otherwise 0. One can then estimate the logit model using standard logistic regression software (Allison 1982, 1995). Many packages also have options for estimating the complementary log-log model.

5. Parametric vs. Semi-parametric Methods

By far, the most popular method for regression analysis of survival data is Cox regression, which combines the proportional hazards model with the partial likelihood method of estimation. Cox regression is sometimes described as *semi-parametric* because, although it is based on a parametric regression model, it does not make specific assumptions about the probability distribution of event times. By contrast, parametric regression models assume particular families of probability distributions, such as exponential, Weibull, Gompertz, lognormal, log-logistic, or gamma.

Although Cox regression is probably the better default method, there are two goals that are easily accomplished with parametric methods but difficult or impossible with Cox regression. First, parametric methods are much better at handling left censoring or interval censoring (especially if the intervals differ across individuals). Second, it is easy to generate predicted times to events with parametric methods, but awkward (and sometimes impossible) to do so with Cox regression. Sometimes people choose parametric methods because they worry that their data do not satisfy the proportional hazards assumption (see Desideratum 15). However, parametric models typically make assumptions that are at least as restrictive as the proportional hazards assumption.

6. Covariates

Issues regarding covariates (also known as predictor variables, independent variables, regressors) are mostly the same in survival analysis as in linear regression and logistic regression (with the important exception described in Desideratum 7). Although it is desirable to provide a rationale for the inclusion of each covariate in the regression model, it is not essential. The consequences of including a variable that actually has no effect are minimal. The real danger, as with any regression analysis of observational data, comes from omitting variables that really have an effect on the outcome. This can lead to severe bias, especially if the omitted variable is moderately to strongly correlated with included variables. So any report of a survival regression should discuss the possibility of important variables that have not been included.

As with other kinds of regression, it is important to consider whether the covariates have nonlinear effects on the outcome and whether there are interactions among the covariates in their effects on the outcome. Strategies for testing and including such nonlinearities and interactions are basically the same as in linear regression, except that there are some special graphical diagnostics available for nonlinearities in Cox regression (Therneau & Grambsch, 2001). Multicollinearity is also a potential problem. Although survival analysis programs typically don't provide collinearity diagnostics, one can simply do a preliminary check with a linear regression program, while specifying the event time as the dependent variable. Because multicollinearity is all about linear relations among the covariates, it is not necessary to evaluate it within the context of a survival analysis.

7. Time-Dependent Covariates

One major difference between survival regression and conventional linear regression is the possibility of time-dependent (time-varying) covariates. These are predictor variables whose values may change over the course of observation. For example, suppose that over a five-year period, information is recorded on any changes in marital status. Then, marital status (updated on a daily basis) could be used as a time-varying predictor of some other event, such as an arrest.

Not all survival analysis methods and/or software can handle time-dependent covariates. For example, most programs for parametric survival models do not allow for time-dependent covariates (although that feature is available in recent releases of Stata). On the other hand, such variables are usually easy to incorporate into discrete-time methods based on logistic (or complementary log-log) regression. That is because each discrete time point is treated as a separate observation, so that any time-dependent covariates can be updated for each observation.

Cox regression is also well known for its ability to handle time-dependent covariates. However, there are two quite different approaches for implementing this capability in software packages. The "episode splitting" method requires that the data be configured so that there is a separate record for each interval of time during which all the covariates remain constant. The "programming statements" method expects one record per individual, with the time-varying covariates appearing as separate variables for each time at which the variables are measured. The time-dependent covariates are then defined in programming statements prior to model specification. Properly implemented, these two methods will give identical results.

One potential issue with time-dependent covariates is that the frequency with which they are measured may not correspond to the precision with which event times are measured. For example, we may know the exact day on which person died of a heart attack. Ideally, a time-dependent covariate, like smoking status, would also be measured on a daily basis. Instead, we may only have annual reports. Some form of imputation is necessary in such cases. The simplest and most common form of imputation is "last value carried forward," although other methods should be considered.

One should also keep in mind that there may be several plausible ways of representing a time-dependent covariate. For example, smoking status could be coded as "person smoked on this day," "number of days out of the last 30 in which the person smoked," or "number of years of smoking prior to the current day," and so forth. Decisions among the alternatives should be carefully considered, and may be based on empirical performance.

8. Repeated Events

If the data contain information on more than one event for each individual, special methods are needed to take advantage of this additional information and to deal with the problems that may arise. If repeated events are observed for an individual, the standard strategy is to reset the clock to 0 each time an event occurs and treat the intervals between events as distinct observations. Thus, if a person is observed to have three arrests over a five-year interval, four observations would be created. The last observation would be a right-censored interval, extending from the third arrest until the end of the observation period.

Repeated events provide more statistical power, and also make it possible to control or adjust for unobservable variables that are constant over time. However, whenever there are multiple observations per individual, there is also likely to be statistical dependence among those observations. Unless some correction is made for this dependence, standard errors and *p*-values will be too low. There are four widely available methods for repeated events that provide appropriate corrections for dependence.

1. Robust standard errors (also known as Huber-White or sandwich estimates) yield accurate standard errors and *p*-values, but leave coefficient estimates unchanged.
2. The method of generalized estimating equations (GEE) also gives corrected standard errors and *p*-values but, in addition, produces more efficient coefficient estimates.
3. Random effects (mixed) models provide the same benefits as GEE, but also correct the coefficients for "heterogeneity shrinkage." This is the tendency of coefficient estimates to be attenuated toward zero because of unobserved heterogeneity.
4. Fixed effects methods also correct for dependence and heterogeneity shrinkage. In addition, they actually control for all stable characteristics of the individual.

For more details, see Allison (1995).

Some of these methods may not be available for some survival regression models or software. For example, Stata will estimate random effects models for Cox regression but SAS will not. Also, note that while fixed effects methods seem to offer the most advantages, they also come with important disadvantages. First, one cannot estimate the effects of variables that are constant over time, like sex or race, although such variables are implicitly controlled. Second, standard errors may be substantially larger because the estimates are based only on variation within individuals.

9. Competing Risks

If a decision has been made to distinguish different kinds of events, an appropriate method must be chosen to handle the different event types. In the competing risks approach, a separate model is specified for the timing of each type of event. These could be any of the models already discussed. If one has continuous time data, each of these models can be estimated separately using standard software for single kinds of events. The trick is that events other than the focal event type are treated as though the individual is censored at that point in time. For example, suppose you want to estimate Cox regression models for job terminations, while distinguishing between quittings and firings. You would estimate one model for quittings, treating firings as censored observations. Then you would estimate a model for firings, treating the quittings as censored observations.

Test statistics are available for testing whether coefficients for a particular variable are the same across event types (Allison, 1995). There are also statistics for testing whether *all* variables have the same coefficients across event types. These statistics can be helpful in determining whether it is really necessary to distinguish the event types. As noted earlier, one disadvantage of distinguishing event types is that the number of events may be small for each event type, leading to a loss of statistical power.

If event times are discrete, maximum likelihood estimation requires that models for competing risks be estimated simultaneously rather than separately. An attractive model that can be estimated with conventional software is the multinomial logit model, also known as the generalized logit model. Unfortunately, there is no comparable multinomial model for the complementary log-log specification.

In some situations with multiple event types, a "conditional" approach may make more sense than competing risks (Allison, 1984). In this approach, the first step is to estimate a model for event timing without distinguishing the different event types. Then, restricting the sample to those individuals who experienced events, the second step is to estimate a binary or multinomial logit model predicting the type of event. This approach is attractive when the event types represent alternative means for achieving a single goal. For example, the event might be the purchase of a computer, and computers are distinguished by whether the operating system is Windows, Linux, or Macintosh.

10. Sampling Issues

There are three questions about sampling that should be addressed: What kind of sample is used? Are the analysis methods appropriate for the sampling method? Is the sample big enough? With regard to the first question, the ideal is a well-designed and executed probability sample. Nevertheless, many survival analyses are carried out on a complete population (e.g., the 50 states in the U.S.) or on convenience samples (e.g., students who volunteered to participate). Although others may disagree, I take the position that survival analysis—including the calculation of confidence intervals and hypothesis tests—is perfectly appropriate for analyzing a complete population. The statistical models that underlie such analyses are based on a hypothesis of inherent randomness in the phenomenon itself, and they do not require any randomization in the study design to justify the application of inferential techniques. The same argument could be made about convenience samples, although any conclusions might only apply to the sample at hand.

Regarding analysis, most survival analysis packages presume, by default, that the sample is a simple random sample. For many samples, however, there will be a need to adjust for clustering, stratification, and/or weighting. Although some packages are explicitly designed for survival analysis with complex samples (e.g., SUDAAN), conventional software can often do the job. Clustering can be accommodated by the methods described above for dependence with repeated events (although it might be difficult to adjust for both repeated events and cluster sampling). Stratification can usually be handled by including the stratification variables as covariates. Finally, most packages allow for differential weighting of observations. However, even if the sampling design involved disproportionate weights, it may not be necessary or desirable to incorporate those weights into the analysis (Winship & Radbill, 1994). This is most likely to be the case if the goal is to estimate an underlying causal model rather than some population regression function.

With regard to sample size, the most important thing to keep in mind is that censored observations contribute much less information than uncensored observations (events). Conventional wisdom has it that there should be at least five (some say 10) events for each parameter in the model, in order for maximum likelihood (or partial likelihood) estimates to have reasonably good properties. As for power considerations, there are numerous software packages and applets that will calculate power and sample size for a single dichotomous covariate. Vaeth and Skovlund (2004) showed how these programs can be easily extended to handle more complex regression problems. Some packages (e.g., Stata, PASS) have routines that will do power calculations for Cox regression analyses.

11. Missing Data

Reports of survival analysis should say something about the extent of missing data and the methods used to handle it. Of course, the default in virtually all survival packages is to do listwise deletion (complete case analysis). And if the proportion of cases lost to missing data is small (say, 10% or less), listwise deletion is probably the best choice. Other conventional methods, like (single) imputation or dummy variable adjustment, typically lead to biased parameter estimates, biased standard error estimates, or both.

For larger fractions of missing data, much better results can be obtained with multiple imputation (Allison, 2001). In this method, imputed values are random draws from the predictive distribution of the missing values given the observed values. Several data sets are created (typically five or more), each with slightly different imputed values. The analysis is performed on each data set using standard software. Then, using a few simple rules, the results are combined into a single set of parameter estimates, standard errors, and test statistics. Multiple imputation uses all the data to produce parameter estimates that are approximately unbiased and efficient. In calculating standard errors and test statistics, multiple imputation, unlike conventional imputation, also incorporates the inherent uncertainty about the values of the missing observations.

Although there are many stand-alone packages for doing multiple imputation, the process is much easier if the imputation is done within the same package used to do the analysis. Software for doing this is available for Stata, SAS, and S-Plus. These also happen to be great packages for survival analysis. Nearly all standard multiple imputation routines are based on the assumption that data are missing at random. This means, roughly, that the probability of missingness may depend on variables that are observed but does not depend on the values of the variables that are missing. Multiple imputation can be done under other assumptions, but the implementation is tricky and must be carefully tailored to each application.

For survival analysis, multiple imputation should only be done for missing values on the predictor variables. Cases that have missing values on the dependent variable should simply be deleted because conventional imputation software is not suited for missing data on event timing and censoring. In setting up the imputation model, however, it is generally a good idea to include both the (logged) event time and the censoring indicator variable so that the relations between these variables and the predictors are adequately reproduced for the imputed variables.

12. Software

Nearly all the major statistical packages have programs for doing Cox regression and Kaplan-Meier estimation of survivor functions. And all can do discrete-time maximum likelihood estimation via logistic regression. Not all can estimate parametric regression models, however, and those that do may vary widely in their capabilities. For example, SAS can estimate parametric models with left and interval censoring but cannot handle time-dependent covariates. With Stata, it is just the reverse. Cox regression programs may also vary widely in their features and capabilities. SPSS, for instance, can handle time-dependent covariates, but its programming functions for defining those covariates are rather limited compared with SAS. As of this writing, I would rate SAS, Stata, and S-Plus as the three best packages for doing survival analysis. Although they vary to some degree in their capabilities, all three have a wide array of programs, functions and options for survival analysis.

Some survival regression programs allow for the incorporation of unobserved heterogeneity into the model. In my judgment, this is a useful feature if individuals have repeated events because it allows for dependence among the multiple observations. However, I would caution against using this option in the more typical case of non-repeated events. In that situation, unobserved heterogeneity models are only weakly identified, and results may depend too critically on the particular specification.

13. Summary Statistics and Data Accessibility

As with other regression methods, it is good practice to report summary statistics for the predictor variables, usually their means and standard deviations. There is a potential complication, however, with time-dependent covariates. If you are using a method that requires multiple records per individual, like discrete-time maximum likelihood or Cox regression using the episode splitting method, you can simply calculate the means and standard deviations over the multiple records. On the other hand, if you are doing Cox regression with programming statements, the time-dependent covariates are created during the estimation process and are not available for calculating descriptive statistics. In that case, I would simply report such statistics for the baseline measurements of the variables.

14. Survivor and Hazard Functions

Although not essential, it is commonplace and informative to present a graph of the survivor function, usually estimated via the Kaplan-Meier method. Such graphs are helpful in giving the reader a sense of the rates of event occurrence and censoring, and how those change over time. In some fields, a cumulative failure graph is preferred over a survivor graph. The two graphs give the same information, however, because the failure probability is just 1 minus the survivor probability.

Even more informative than the survivor function is a graph of the estimated hazard function because it more directly quantifies the rate of event occurrence and how that rate changes over time. But the problem with the hazard function is that non-parametric estimates based on Kaplan-Meier require smoothing, and different smoothing algorithms can yield markedly different graphs. Therefore, if hazard graphs are to be presented, I recommend using the actuarial (life table) method. Although this requires an arbitrary choice of time intervals, results tend to be more stable than those produced by smoothing methods.

15. Proportional Hazards Assumption

Cox regression is based on the proportional hazards model. The proportional hazards assumption says, in essence, that the dependence of the hazard on time has the same basic shape for everyone, even as the magnitude of the hazard varies across individuals as a function of their predictor values. A crucial implication of this assumption is that predictor variables have the same effects at all points in time, that is, there are no interactions with time.

Although many researchers get very concerned about whether their data satisfy this assumption, I believe that those concerns are often unwarranted. If the assumption is violated for a particular predictor variable, it simply means that the coefficient for this variable represents a kind of "average" effect over the period of observation. For many applications, this may be sufficient. In some cases, however, the violations may be so severe that they lead to biases in the effects of other variables. In other cases, there may be direct interest in how the effect of a variable on the hazard changes over time.

A quick check of the proportional hazards assumption can be obtained by computing correlations between time (or some function of time) and "Schoenfeld residuals" which are calculated separately for each predictor. Non-zero correlations are evidence against the proportionality assumption. Several Cox regression software packages have an option to compute these statistics.

A more definitive check is to directly include interactions between predictors and time, which are specified as time-dependent covariates. Significant interactions indicate violation of the assumption. However, in this case the method of diagnosis is also the cure. By including the interactions, the Cox model is extended to allow for non-proportional hazards.

Another way to allow for non-proportional hazards is the method of stratification, which allows for different hazard functions for different categories of a categorical variable (like sex or marital status). This is a good method for controlling for a variable without imposing the proportional hazards assumption. But it does not yield any estimates of the effect of that variable, nor does it give a test of the proportional hazards assumption.

16. Model Comparisons

Researchers typically want to know how well their statistical models fit the data. Unfortunately, global or absolute measures of fit are generally not available for survival analysis models. Usually, the best we can do is to compare the relative fit of different models. If the models are nested (i.e., one model can be obtained from another by imposing restrictions on the parameters), likelihood ratio tests can be calculated by taking twice the positive difference in the log-likelihoods for the two models. Such tests can tell you whether the more complicated model is significantly "better" than the simpler model. These tests are especially useful when estimating parametric models because some of the better-known parametric distributions are nested within the generalized gamma distribution.

If two models are not nested, informal comparisons can be accomplished with Akaike's information criterion (AIC) or Schwarz's Bayesian Information Criterion (SBC or BIC). These statistics "penalize" the log-likelihood for the number of covariates in the model, enabling one to validly compare models with different sets of covariates. Many software packages report one or both of these statistics, both for parametric models and for Cox regression models. Preference is given to models with lower values of these statistics, although no p-values can be calculated.

17. Reports of Coefficients and Associated Statistics

Results for Cox regression may be reported as either beta (β) coefficients or hazard ratios (a few authors report both). Beta coefficients are more easily interpreted with respect to sign (positive, negative, or zero). However, their numerical magnitudes are difficult to interpret. Hazard ratios (which are always positive) may confuse some readers because a value of 1 means no effect. But the numerical magnitude has a more straightforward meaning: if HR denotes the hazard ratio, 100(HR–1)% is the percentage change in the hazard for a one-unit increase in the predictor. In this respect, they behave just like odds ratios in logistic regression. In the biomedical sciences, there is a clear preference for reporting hazard ratios, and this preference seems to be spreading to other fields as well.

If you report β coefficients, you should also report either standard errors or 95% confidence intervals. Because hazard ratios have asymmetric distributions, standard errors are not generally reported. Instead, the convention is to report 95% confidence intervals. It is optional but desirable to report p-values for testing the null hypothesis of no effect for each coefficient. Also desirable is a chi-square test for the null hypothesis that all coefficients are zero. Many authors ritualistically report the log-likelihood for each model, but this is usually not informative (unless it can be used to compare nested models).

18. Conditional Survivor or Hazard Functions

In Desideratum 14, I discussed the use of survivor or hazard functions as a descriptive device. After estimating a regression model, it is often desirable to illustrate its implications by displaying a model-based survivor function or hazard function. For example, if interest centers on the effect of some treatment, one could plot survivor functions for the treated vs. control groups in such a way that the plots embody any model assumptions (e.g., proportional hazards) and also control for any covariates in the model. If the variable of interest is quantitative, one can produce plots for several selected values of that variable, again while adjusting for any covariates.

19. Potential Methodological Limitations

Any application of statistical methods to real-world data is vulnerable to errors of one sort or another. Researchers need to be acutely aware of potential problems with their data and with the analytic methods they apply to those data. They also need to be upfront with their readers regarding any problems that they suspect could compromise their conclusions.

As noted in Desideratum 6, the most serious potential problem with survival analysis regression methods is the same as that for any other regression method applied to observational (non-experimental) data: the omission of variables (confounders) that affect the outcome and that are also correlated with the included variables. The omission of confounders can produce biases so severe that they lead to conclusions that are the exact opposite of the true state of affairs.

A problem peculiar to survival analysis is informative censoring (see Desideratum 3). Once the data are in hand, there is not much that can be done about this. But, if the number of randomly censored cases is substantial, research reports should discuss their potential impact. A sensitivity analysis can help to discern the potential direction of biases resulting from informative censoring.

Another potential danger comes from fitting an incorrect model. Some of the comparative statistics discussed in Desideratum 16 can be helpful in finding a good model. But it is also desirable to fit rather different models to the data and see if the results are consistent across models. For example, there is no good way to compare the fit of a Cox regression model with parametric gamma model. But it can be quite useful to fit both models to see if they lead to the same conclusions. If they do, well and good. If not, then your confidence in the results should be appropriately reduced.

References

Allison, P. D. (1982). Discrete time methods for the analysis of event histories. In S. Leinhardt (Ed.), *Sociological methodology 1982* (pp. 61–98). San Francisco: Jossey-Bass.

Allison, P. D. (1984). *Event history analysis.* Thousand Oaks, CA: Sage.

Allison, P. D. (1995). *Survival analysis using SAS: A practical guide.* Cary, NC: SAS Institute.

Allison, P. D. (2001). *Missing data.* Thousand Oaks, CA: Sage.

Blossfeld, H.-P., & Rohwer, G. (2001) *Techniques of event history modeling: New approaches to causal analysis.* Mahwah, NJ: Erlbaum.

Box-Steffensmeier, J., & Jones, B. (2004). *Event history modeling: A guide for social scientists.* Cambridge, UK: Cambridge University Press.

Collett, D. (2003). *Modeling survival data in medical research.* New York: Chapman & Hall.

Hosmer, D. W., & Lemeshow, S. (2003). *Applied survival analysis: Regression modeling of time to event data.* New York: Wiley.

Kiefer, N. M. (1988). Economic duration data and hazard functions. *Journal of Economic Literature*, *26*, 646–679.

Klein, J. P., & Moeschberger, M. L. (2003). *Survival analysis: Techniques for censored and truncated data.* New York: Springer-Verlag.

Kleinbaum, D. G., & Klein, M. (2005). *Survival analysis: A self-learning text.* New York: Springer-Verlag.

Teachman, J. D. (1983). Analyzing social processes: Life tables and proportional hazards models. *Social Science Research*, *12*, 263–301.

Therneau, T. M., & Grambsch, P. M. (2001). *Modeling survival data: Extending the Cox model.* New York: Springer-Verlag.

Tuma, N. B., & Hannan, M. T. (1984). *Social dynamics: Models and methods.* Orlando, FL: Academic Press.

Væth, M., & Skovlund, E. (2004). A simple approach to power calculations in regression models. *Statistics in Medicine*, *23*, 1781–1792.

Winship, C., & Radbill, L. (1994). Sampling weights and regression analysis. *Sociological Methods & Research*, *23*, 230–257.