■ SPECIAL ARTICLE

OPEN
CME

Survival Analysis and Interpretation of Time-to-Event Data: The Tortoise and the Hare

Patrick Schober, MD, PhD, MMedStat,* and Thomas R. Vetter, MD, MPH†

Survival analysis, or more generally, time-to-event analysis, refers to a set of methods for analyzing the length of time until the occurrence of a well-defined end point of interest. A unique feature of survival data is that typically not all patients experience the event (eg, death) by the end of the observation period, so the actual survival times for some patients are unknown. This phenomenon, referred to as censoring, must be accounted for in the analysis to allow for valid inferences. Moreover, survival times are usually skewed, limiting the usefulness of analysis methods that assume a normal data distribution. As part of the ongoing series in *Anesthesia & Analgesia*, this tutorial reviews statistical methods for the appropriate analysis of time-to-event data, including nonparametric and semiparametric methods—specifically the Kaplan-Meier estimator, log-rank test, and Cox proportional hazards model. These methods are by far the most commonly used techniques for such data in medical literature. Illustrative examples from studies published in *Anesthesia & Analgesia* demonstrate how these techniques are used in practice. Full parametric models and models to deal with special circumstances, such as recurrent events models, competing risks models, and frailty models, are briefly discussed.   (Anesth Analg 2018;127:792–8)

The surprising thing about young fools is how many survive to become old fools.

—Doug Larson (1926–2017), American journalist, columnist, and editor

The occurrence of a well-defined event such as patient mortality is often a primary outcome in medical research. This is essentially a binary outcome (the event has occurred versus it has not occurred). In a previous tutorial in this series, we described how such binary outcome data can be analyzed with logistic regression.[1] For example, one can estimate the relationship between one or more covariates, also referred to as independent variables or predictor variables (eg, treatments or prognostic factors) and the odds of experiencing the outcome within a specific time frame (eg, mortality within 30 days postoperatively).

However, logistic regression analysis is not appropriate when the research question involves the length of time until the end point occurs—for example, estimating median survival times, plotting survival over time after treatment, or estimating the probability of surviving beyond a prespecified time interval (eg, 5-year survival rate). Researchers are also often interested in whether survival times are related to covariates, and estimating the effect size of a specific covariate (eg, magnitude of the treatment effect) when it is adjusted for potential confounders.

Furthermore, it may initially appear that such a research question about the length of a time interval, which is essentially a continuous outcome variable, can be addressed by linear regression or related techniques like a *t* test or analysis of variance.[1,2] However, a key distinction between survival times and other continuous data is that the event of interest (eg, death) will usually have occurred only in some but not in all patients by the time the study ends.

For patients who survive until the end of the study period, or who are lost to follow-up before the end of the observation period, full survival times are unknown. Instead all that is known is that the survival time is greater than the observation time. This unique feature of survival data is referred to as right censoring, which is described in more detail below.[3]

Ignoring censored patients in the analysis, or simply equating their observed survival time (follow-up time) with the unobserved total survival time, would bias the results. Even if there was no censoring in the data set, survival times usually have a heavily skewed distribution, limiting the usefulness of statistical tests that assume a normal data distribution.[3]

Analyzing survival data is unique in that the research interest is typically a combination of whether the event has occurred (binary outcome) and when it has occurred (continuous outcome). Appropriate analysis of survival data requires specific statistical methods that can deal with censored data. As the assessed outcome is frequently mortality, these techniques are subsumed under the term survival analysis.

More generally, however, these techniques can be used for the analysis of the time until any event of interest occurs (eg, recurrence of a disease; initial, breakthrough postoperative pain; or failure of an implanted medical device), and such data can thus also be called time-to-event or failure

time data.[4,5] In this tutorial, we use the terms survival time, time-to-event, and failure time synonymously.

As part of the ongoing series in *Anesthesia & Analgesia*, this basic tutorial reviews statistical methods that are appropriate for survival data. We focus on the most common techniques, which are the Kaplan-Meier estimator, log-rank test, and the Cox proportional hazards (PH) model. The Table provides a summary of key-related terms. An extensive discussion of full parametric techniques and the special circumstances that call for other techniques is beyond the scope of this basic tutorial, and we thus intentionally provide only precursory coverage of these more advanced aspects.

CONCEPTS AND TERMINOLOGY IN SURVIVAL ANALYSIS

General Considerations

The event of interest should be clinically relevant, well defined, unambiguous, and preferably easily observable. While the patient death seems to be such an unambiguous end point, misclassification is possible when specific-cause mortality, rather than all-cause mortality, is the outcome of interest.[3] Some end points, such as recurrence of cancer, have the disadvantage that they do not occur instantaneously, making it difficult to specify the exact time point of occurrence. In such settings, the clearest description of the outcome is often "time-to-detection" rather than "time-to-event." Whatever the event of interest, a clear and unambiguous definition is essential.

The total length of follow-up and follow-up intervals should be sensibly chosen to ensure that a sufficient number of events are observed (see Power and Sample Size Considerations section) and that the timing of occurrence can be determined, to avoid interval censoring (as detailed below). Conversely, long-term observational studies carry the risk that factors that influence survival time, other than the treatment or factor under investigation, may also change during the study period. Patients recruited to the study early should ideally have the same risk of event occurrence as patients recruited late.[3]

As the failure time is the time between some starting point (origin) and the event, not only the event but also the time of origin needs to be clearly specified. Ideally, the time of origin should also be sensibly chosen, so all individuals are as much as possible on a par.[6] In a study comparing therapeutic interventions on a survival outcome, the starting point is typically the time when the intervention is administered.

Table. Basic Terminology and Definitions

Term	Definition
Censoring	Incomplete observation of survival times; subjects with an incomplete observation are referred to as censored.
Left censoring	The patient is known to have experienced the event before the start of the observation period, so the actual time-to-event is shorter than the interval between the origin and start of observation, but it is unknown by how much.
Right censoring	Observation of the patient is terminated before the event occurs, so the actual time-to-event, if it were to occur, is longer than the observation time, but it is unknown by how much.
Cox regression	A semiparametric technique that is commonly used for survival data analysis that allows one to simultaneously assess the association between multiple covariates and survival, also termed Cox proportional hazards model.
Delayed entry	Synonym for left truncation.
Failure rate	Synonym for hazard rate.
Failure time	Time until an event occurs (interval between origin and event); synonym for survival time or time-to-event.
Hazard function	Describes the instantaneous rate of occurrence of the event over time.
Hazard rate	Rate of occurrence of the event during some time interval; often used to denote the instantaneous rate in an infinitesimally small time interval.
Hazard ratio	Ratio of 2 hazard rates (eg, for 2 different treatment groups).
Kaplan-Meier curve	Graphical representation of the survival function with probabilities estimated by the Kaplan-Meier method.
Kaplan-Meier method	Nonparametric method to estimate survival probabilities over time.
Lead time	The time between the early diagnosis of a specific disease with screening and the time in which diagnosis would have been made without screening.
Lead-time bias	Lead time increases the observed survival time even if there was no effect of screening on absolute survival time; hence, it biases estimates of survival time.
Log-rank test	Nonparametric hypothesis test to compare 2 or more survival functions.
Nonparametric method	Statistical method that neither imposes assumptions on the distribution of the outcome variable (a specific shape of the survival function or hazard function) nor assumes a functional form of the covariates.
Origin	Starting point in the timeline of a survival time, which is often the time point when a disease or condition is diagnosed, or the time point of a specific event (eg, birth or operation).
Parametric method	Statistical method that assumes a specific distribution of the outcome variable, including the relationship between covariates and survival.
Proportional hazards model	Used to model the hazard function, assuming that the effect of a covariate is the same at all time points; such models can be either semiparametric (Cox regression model) or parametric.
Semiparametric method	Statistical method that does not assume a specific distribution of the outcome variable (the survival function or hazard function), but does assume a specific relationship between covariates and the outcome.
Survival data	Umbrella term covering data that describe the time until any well-defined event (not necessarily death) occurs.
Survival (survivor) function	Describes the probability over time that the event of interest has not yet occurred.
Survival time	Time until an event occurs (interval between origin and event, which does not necessarily have to be the death of a subject), synonym for failure time, or time-to-event.
Time-to-event	Time until an event occurs (interval between origin and event), synonym for survival time or failure time.
Truncation	Subject selection depending on whether or not the event has occurred.
Left truncation	Selective inclusion of patients in whom the event has not occurred early; patients who have already experienced the event before the time point of patient identification are not identified and often may not even be known to exist.
Right truncation	Selective inclusion of patients in whom the event has occurred (eg, due to selection of patients from a death registry).

In epidemiologic or screening studies, the origin is often when a condition or disease is diagnosed. However, this can lead to biased estimates of survival times, especially when the study intervention does not only presumably affect the event but also the time of origin. This so-called lead-time bias is common in studying whether a screening program for a specific disease increases survival time. Observed prolonged survival in screened patients could be merely due to the disease being diagnosed earlier, and not necessarily reflect a benefit of screening on absolute survival time.[7]

Truncation

As in any clinical study, the target population, as well as inclusion and exclusion criteria, must be clearly defined. A unique feature of survival data is truncation, which results from selection bias and refers to subject selection depending on whether or not the event has occurred.[8] Subjects may only be identified for observation at some time point after their respective time of origin.[9]

Patients who have already experienced the event before the time point of patient identification may not be identified—for example, because they have already died, these patients may not be known to exist.[8] In this situation, only those patients who have not experienced the event will selectively enter the study, which is referred to as left truncation or delayed entry.

Alternatively, right truncation occurs when patients who experience the event are selectively included[6]—for example, when patients are obtained from a death registry, and hence, survivors are not selected for the study. While the bias caused by truncation can be partially addressed during the analysis, it is often preferable be prevent it at the design phase of the study.

Censoring

Censoring refers to incompletely observed survival times and is inherent with most survival data. The situation described above in which not all the patients experience the event until completion of the study is referred to as right censoring. Visualizing the timeline of a patient's observed survival time, the unobserved event, if it were to occur, would lie beyond the right side of the time point at which the patient is censored. Right censoring is the most common type of censoring in survival studies, and the statistical methods described below are well suited to deal with this type of censoring. Basically, censored patients are: (1) included in estimates of survival probabilities at time points preceding their censoring time point; and (2) excluded from the analysis thereafter.[10]

Unbiased inferences require the censoring to be noninformative, with the time of censoring absolutely not related to the event time.[3] Informative censoring would occur when patients are censored due to a medical condition that is related to the future risk of the event (eg, inability to show up for a clinic visit due to severe illness and thus loss to follow-up). Unfortunately, this problem is neither easy to detect nor is there an ideal solution—other than conducting the study in a way that promotes complete follow-up and avoids informative censoring.[11] If possible, data on the reasons for loss-to-follow-up should be collected because such information can be used in sensitivity analyses to assess for potential bias.

Left censoring occurs when a subject is known to have had the event before the start of the observation, but the exact time of the event is unknown. This contrasts with left truncation, where the patient is often not even known to exist.[12] Similarly, interval censoring is where it is only known that the event occurred between 2 time points, but again, the exact time is unknown. Left and interval censored data are less common and usually do not exist when death is the outcome of interest. Statistical techniques to deal with left and interval censored data are available; however, they are infrequently used and will not be covered in this basic tutorial.

Survival (Survivor) Function, Hazard Rate, Hazard Function, and Hazard Ratio

The survival (or survivor) function and the hazard function are fundamental to survival analysis. The survival function describes the probability of surviving past a specified time point, or more generally, the probability that the event of interest has not yet occurred by this time point (Figure 1).[13]

A hazard rate (or failure rate) is the rate of occurrence of the event during a given time interval.[10] The hazard function describes the instantaneous rate of occurrence over time, which can conceptually be viewed as the hazard rate during an infinitesimally small time interval. The hazard and survival functions are closely related and can easily be converted to each other.[3] When the hazard rate is high, survival declines rapidly and vice versa.

While it is not necessary to understand the hazard function in detail, it is the basis of PH models, which are extensively used to model survival data. Importantly, the exponentiated parameter estimates of these models can be interpreted as a hazard ratio (HR), which is an estimate of the ratio of the hazard rates between 2 groups (eg, treatment versus control).[14]

The HR is similar to the risk ratio (relative risk), with a value higher or lower than 1 indicating a higher or lower hazard rate, respectively, than the comparison group. While the HR is technically not the same as the risk ratio, it is often conveniently interpreted as such in the literature.[15]

General Overview of Methods to Analyze Survival Data

In analyzing survival data, 3 common classes of methods are broadly distinguished:

1. Nonparametric methods, which neither impose assumptions on the distribution of survival times (a specific shape of the survival function or hazard function) nor assume a specific relationship between covariates and the survival time. This class includes the Kaplan-Meier estimator and log-rank test.
2. Semiparametric methods also make no assumptions regarding the distribution of survival times but do assume a specific relationship between covariates and the hazard function—and hence, the survival time. The widely used Cox PH model is a semiparametric method.
3. Parametric methods assume a distribution of the survival times and a functional form of the covariates.

Figure 1. Survival (survivor) function estimated by the Kaplan-Meier method, including 95% confidence bands. Censoring is indicated by vertical marks (at 5 and 21 d). The number of patients at risk at different time points is displayed on the graph. The point on the x-axis where the horizontal dashed line at a survival probability of .5 intersects the curve represents the estimated median survival time (17 d).
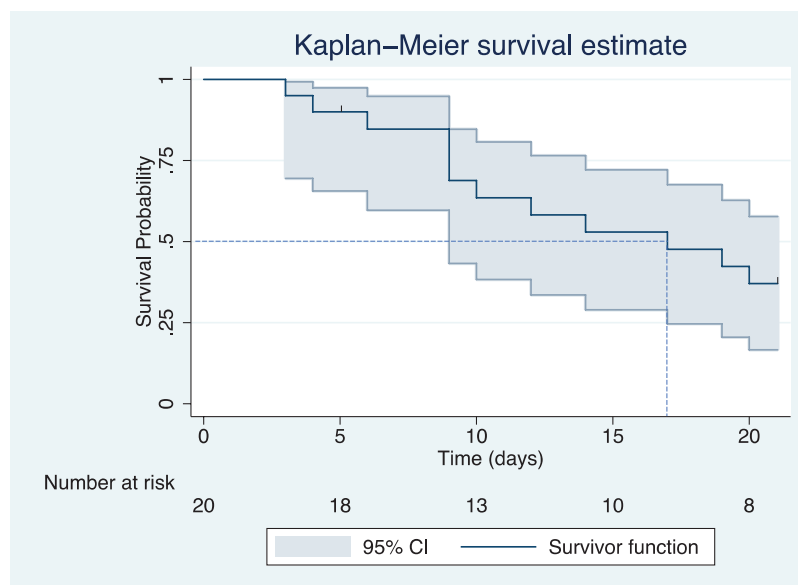
KAPLAN-MEIER ESTIMATOR

The Kaplan-Meier method estimates the unadjusted probability of surviving beyond a certain time point.[16] A Kaplan-Meier curve shows the estimated survival function by plotting estimated survival probabilities against time (Figure 1).[3] The estimated survival probability is constant between the events. Therefore, the curve is a step-function in which each vertical drop indicates the occurrence of one or more events.[17] Right censoring of patients is typically indicated by a vertical mark at the censoring time or other symbols like an asterisk.[6]

CIs for the survival probabilities can be readily calculated, and confidence bands can be plotted around the survival function (Figure 1).[3] These CIs provide an estimate of the range of plausible values of the survival probability in the population from which the patients were sampled.[18] Often, several Kaplan-Meier curves of different groups (eg, treatment groups or prognostic factors) are plotted together in 1 graph, allowing for a visual comparison of the survival probabilities (Figure 2).

The median survival time, which is the time by when the event has occurred in 50% of the patients or study subjects, is a commonly reported summary statistic for survival time data. This median survival time can be conveniently estimated from the Kaplan-Meier curve as the x-axis (time) value at the point where an (imaginary) horizontal line at the 50% survival probability on the y-axis crosses the survival curve (Figure 1).[10] Additionally or alternatively, the survival probability at appropriate time points (eg, at 1 and 5 years) can be reported.

LOG-RANK TEST

Analogous to comparing groups of continuous data using a t test or analysis of variance, the survival curves for 2 or more different groups (eg, treatments or prognostic factors) can also be compared with hypothesis testing. Most commonly, the log-rank test is applied, which tests the null hypothesis that there is no difference in the probability of an event at any time point.[19]

When reporting a log-rank P value comparing Kaplan-Meier curves, researchers should make clear that the entire distribution is being tested, and not a particular time such as 5-year survival. The log-rank test is based on the same assumptions as the Kaplan-Meier survival curve, and makes no explicit assumptions about the distribution of the survival curves.

However, when the survival curves of different groups cross—indicating that 1 group has a more favorable survival in a certain time interval and less favorable survival in another time interval—the power to detect such differences is very low.[19] Moreover, the log-rank test cannot adjust for other covariates that might affect survival time. While it can determine whether observed differences are significant, it cannot provide an estimate of the difference between groups.[14] Other techniques, described below, can be used to address these issues.

COX PH MODEL

The Cox PH model is the most commonly used survival data analysis technique that simultaneously allows one to include and to assess the effect of multiple covariates.[14] These model covariates can include the variables of specific research interest (treatment groups), as well as potential confounders for which the researcher wants to control (demographic and other clinical factors). Multiple strategies for covariate selection have been described, and the aim of the study—most often to determine the effect of a covariate while controlling for confounding versus prediction of survival using a set of predictor variables—should be considered in choosing a strategy.[11,20,21]

Cox PH regression actually does not directly model survival probabilities or survival times, but the hazard function.[22] Herein, it is assumed that all patients have a common baseline hazard function that only depends on time. Each subject's individual hazard function is a multiple of this common baseline hazard, and the individual multiplicator is a constant, determined by a time-independent function of a patient's individual covariate values.[22] This implies that
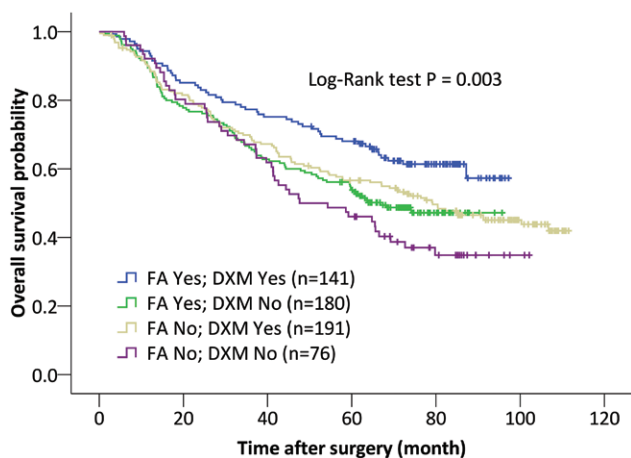
Figure 2. Kaplan-Meier curves displaying the estimated survival probability for 4 different groups of patients after lung cancer surgery. Patients either did or did not perioperatively receive flurbiprofen axetil (FA) and dexamethasone (DXM) (reprinted with permission from Huang et al[29]). Each vertical step in the curve indicates one or more events (ie, deaths), and right-censored patients are indicated by a vertical mark in the curve at the censoring time. A visual inspection suggests that survival seems to be more favorable for patients who received FA and DXM, compared with patients who received none of these 2 drugs. The log-rank test indicates a significant difference between the survival curves.

the ratio of the hazard rates between different patients (the HR) is assumed to be constant over time—in other words, the effect of a covariate is assumed to be the same at all time points. This is the PH assumption of the Cox PH model,[6] which is discussed in more detail below.

Under this assumption, exponentiated regression coefficients for each covariate can be interpreted in terms of the HR for a 1-unit change in the respective covariate value. This is akin to the interpretation of exponentiated regression coefficients as odds ratios in logistic regression.[23]

While the Cox PH model estimates regression coefficients without making assumptions about the shape of the hazard function, it is possible to work backward and use the parameter estimates to estimate the adjusted hazard or adjusted survival function. This allows the plotting of adjusted curves for different groups, which are very similar to Kaplan-Meier curves, but instead show or predict the probability of survival in each group, while keeping the other covariates fixed at their mean values.[24] Survival proportions can also be predicted for each arbitrary combination of covariate values.[14]

Model Assumptions

Assumptions of the Cox PH model—other than assumptions that apply to all survival analyses, such as noninformative censoring described above—include a linear relationship between the covariates and the log-hazard, as well as the PH assumption.

While the PH assumption is central to the Cox model, its actual importance is debated. While some authors stress the importance,[25,26] others de-emphasize it and take the view that the HR can still be viewed as an average effect during the observation period when the assumption is violated.[27] We recommend that the PH assumption should be assessed because it is sometimes possible to fix violations. Several

methods to check this and other assumptions have been suggested, and we refer to previous literature on the topic for a detailed overview.[21,25]

The Cox PH model assumes the covariates to be time-independent—in other words, the values of the variable of each patient (eg, gender and age at time of diagnosis) do not change over time. Extensions of the Cox model are available that allow for covariates that vary over time (eg, blood pressure recordings at follow-up time points).[11] However, such models should be used carefully, as they are difficult to interpret, prone to misspecifications, and markedly increase the potential for erroneous inferences.[28]

PUBLISHED REPRESENTATIVE USE OF NONPARAMETRIC AND SEMIPARAMETRIC TECHNIQUES

Nonparametric and semiparametric methods are commonly used to analyze survival data in anesthesia, critical care, perioperative, and pain research.[29–35] We illustrate the practical use of these techniques in 3 different types of studies recently published in *Anesthesia & Analgesia*.[29–31]

In their retrospective cohort study, Huang et al[29] sought to identify predictors of long-term survival in patients after lung cancer surgery. The authors initially used multiple log-rank tests to identify covariates that are potentially related to survival. Those covariates that were either considered clinically important or displayed a P value <.2 in their log-rank test were entered in a multivariable Cox PH model. With this model, the authors identified 6 factors associated with either longer or shorter overall survival. For example, limited resection was associated with a higher hazard rate and hence shorter survival (HR, 1.46; 95% CI, 1.08–1.98; P = .013), whereas perioperative use of dexamethasone was associated with prolonged survival (HR, 0.70; 95% CI, 0.54–0.90; P = .006).[29] The authors also specifically compared patients who received dexamethasone and flurbiprofen axetil with patients who received only one of either drugs or none of the drugs. Kaplan-Meier curves were presented for each of the 4 possible combinations (Figure 2), and a log-rank test was used for an unadjusted comparison of the survival curves. A multivariable model adjusting for confounders suggested that administration of both flurbiprofen axetil and dexamethasone was associated with prolonged overall survival when compared to no use of both, with an adjusted HR of 0.57 (95% CI, 0.38–0.84; P = .005).[29]

In their randomized controlled trial, Wilson et al[31] studied whether dural puncture epidural (DPE)—a technique where the dura is punctured but medication is not administered in the subarachnoid space—expedites analgesia in laboring patients compared to the conventional lumbar epidural (LE) technique. Adequate analgesia was defined as a score of ≤10 mm on a 100-mm Visual Analog Scale during active contractions. The authors applied the Kaplan-Meier method to estimate median time to achieve adequate analgesia in each treatment group. A Cox PH regression model with treatment group as a sole independent variable was used to estimate the treatment effect. Median time to adequate analgesia was 8 minutes (95% CI, 6–10 minutes) in the DPE group and 10 minutes (95% CI, 8–14 minutes) in LE group. The estimated HR was 1.67 (95% CI, 1.02–2.64; P = .042).[31] To make these results more understandable, the

authors followed the common (albeit not statistically correct) practice of interpreting the HR as a relative risk: "The relative risk of achieving pain control at any collected time point in parturients receiving DPE was 1.7 times greater than in those receiving LE."[31] Of note, in the above first example, observational data were analyzed using the Cox PH model to adjust for confounding. Here, in this randomized controlled trial, the purpose of the Cox PH model was to obtain an estimate of the treatment effect.

Using data on patients who participated in 2 trials across 4 clinical sites for a follow-up analysis, Podolyak et al[30] studied effects of supplemental perioperative oxygen on long-term mortality in patients undergoing colorectal surgery. In the 2 original trials, patients had randomly received either 30% or 80% inspired oxygen perioperatively. The authors present survival curves using Kaplan-Meier estimates and use a Cox PH model, stratified by study and site to allow for separate baseline hazards for each study and site. This approach was (presumably) chosen as it allows for the estimation of an overall HR estimate and significance test across all study sites. No effect of 80% vs 30% inspired oxygen was observed on mortality, with an overall estimated HR of 0.93 (95% CI, 0.72–1.20; P = .57).[30]

PARAMETRIC MODELS

Parametric models assume a specific distribution of the survival times. Advantages of a parametric model include a higher efficiency (ie, greater power),[14] which can be particularly useful with smaller sample sizes. Furthermore, a variety of parametric techniques can model survival times when the PH assumption is not met.

However, it can be quite challenging to identify the most appropriate data distribution, and parametric models have the drawback of providing misleading inferences if the distributional assumptions are not met. In contrast, the semi-parametric Cox model is a safe and proven method without the need to specify a specific data distribution,[36] which is why this model is most common in analyzing survival data. For a more detailed discussion on parametric models, we refer to previously published literature on the topic.[14,36]

RECURRENT EVENTS, COMPETING RISKS, AND FRAILTY MODELS

The previously described techniques are useful for studying time until occurrence of a specific event that occurs only once, terminates the observation of a patient, and occurs independently between the patients. While this situation is common in many time-to-event study designs, researchers may be interested in: (1) events that can occur more than once or in a series of events in which each event has its own failure time; (2) situations in which follow-up may be terminated by >1 event; or (3) clusters of patients for whom the event does not occur independently.

Recurrent event models are capable of modeling the sequential occurrence of events over time.[37] This can be the same event transpiring several times (eg, occurrence of myocardial infarction) or a series of different ordered events (eg, different, progressive stages of a disease until death occurs).

Competing risk models can accommodate multiple (competing) types of failure events, each of which terminate the observation of an individual.[38] For example, a researcher may not want to simply consider all-cause mortality as the failure event, but would like to study the relationship between covariates and specific-cause mortality. Or, commonly, researchers are interested in an event such as cancer recurrence, but death that occurs before the event of interest is a competing risk. In this setting, the researcher can either model the time to the earliest of death or cancer recurrence or use special methods to model both events.

Frailty models account for nonindependence of observations in clustered data (for correlated failure times), by incorporating random effects.[39] Such data may arise when the survival times of individuals within a cluster (eg, family or hospital) tend to be more similar to each other than survival times of patients who belong to different clusters. These models are analogous to mixed effect models for uncensored longitudinal and correlated data, as described in a recent tutorial in this series.[40]

POWER AND SAMPLE SIZE CONSIDERATIONS

The power of a method to analyze survival time data depends on the number of events rather the total sample size.[21] Therefore, calculation of total sample size is a 2-step process.[41]

First, the number of events needed to detect a minimum clinically important effect size, like a prespecified HR, with a preselected power and alpha level is computed. Depending on the planned data analysis method, different approaches for estimating the number of events have been proposed, including the Schoenfeld method for log-rank tests or PH models.[42]

Second, to calculate the total sample size, the proportion of patients who are expected to experience the event needs to be estimated.[41] Of note, for multivariable models like the Cox PH model, it has been suggested that at least 10 events need to be observed per covariate to be included in the model.[43]

CONCLUSIONS

Survival data are unique in that the research questions essentially involve a combination of whether the event has occurred in the observation period and when it has occurred. Censoring, or the incomplete observation of failure times, is common in these data, such that specific statistical methods are required for an appropriate analysis.

The Kaplan-Meier method estimates the unadjusted probability of surviving beyond a certain time point, and a Kaplan-Meier curve is a useful graphical tool to display the estimated survival function. The log-rank test is commonly used to compare survival curves between different groups, but can only be used for a crude, unadjusted comparison.

The Cox PH model is the most commonly used technique to assess the effect of factors, such as treatments, that simultaneously allows one to control for the effects of other covariates. The exponentiated regression coefficients can be interpreted in terms of an HR. This semiparametric technique makes no assumptions about the distribution of the survival times.

If the distribution can be appropriately identified and modeled, parametric techniques can alternatively be used. For special circumstances in which the standard techniques cannot be validly used, a variety of methods including

recurrent events models, competing risks models, and frailty models are available. ■■

DISCLOSURES

Name: Patrick Schober, MD, PhD, MMedStat.
Contribution: This author helped write and revise the manuscript.
Name: Thomas R. Vetter, MD, MPH.
Contribution: This author helped write and revise the manuscript.
This manuscript was handled by: Jean-Francois Pittet, MD.

REFERENCES

1. Vetter TR, Schober P. Regression: the apple does not fall far from the tree. *Anesth Analg*. 2018;127:277–283.
2. Vetter TR, Mascha EJ. Unadjusted bivariate two-group comparisons: when simpler is better. *Anesth Analg*. 2018;126:338–342.
3. Clark TG, Bradburn MJ, Love SB, Altman DG. Survival analysis part I: basic concepts and first analyses. *Br J Cancer*. 2003;89:232–238.
4. Kasza J, Wraith D, Lamb K, Wolfe R. Survival analysis of time-to-event data in respiratory health research studies. *Respirology*. 2014;19:483–492.
5. Kalbfleisch JD, Prentice RL. *The Statistical Analysis of Failure Time Data*. 2nd ed. Hoboken, NJ: John Wiley & Sons; 2002.
6. Kartsonaki C. Survival analysis. *Diagn Histopathol*. 2016;22:263–270.
7. Baum M. Breast cancer screening comes full circle. *J Natl Cancer Inst*. 2004;96:1490–1491.
8. Dai H, Wang H. Introduction. In: *Analysis for Time-to-Event Data Under Censoring and Truncation*. London, UK: Academic Press Elsevier Ltd, 2017:1–13.
9. Tsai WY, Jewell NP, Wang MC. A note on the product-limit estimator under right censoring and left truncation. *Biometrika*. 1987;74:883–886.
10. Blagoev KB, Wilkerson J, Fojo T. Hazard ratios in cancer clinical trials–a primer. *Nat Rev Clin Oncol*. 2012;9:178–183.
11. Clark TG, Bradburn MJ, Love SB, Altman DG. Survival analysis part IV: further concepts and methods in survival analysis. *Br J Cancer*. 2003;89:781–786.
12. Cain KC, Harlow SD, Little RJ, et al. Bias due to left truncation and left censoring in longitudinal studies of developmental and disease processes. *Am J Epidemiol*. 2011;173:1078–1084.
13. Kleinbaum DG, Klein M. Introduction to survival analysis. In: *Survival Analysis*. 3rd ed. New York, NY: Springer, 2012:1–54.
14. Bradburn MJ, Clark TG, Love SB, Altman DG. Survival analysis part II: multivariate data analysis–an introduction to concepts and methods. *Br J Cancer*. 2003;89:431–436.
15. Kalbfleisch JD, Prentice RL. Relative risk (Cox) regression models. In: *The Statistical Analysis of Failure Time Data*. 2nd ed. Hoboken, NJ: John Wiley & Sons, 2002:95–147.
16. Hosmer DW, Lemeshow S, May S. Descriptive methods for survival data. In: *Applied Survival Analysis*. 2nd ed. Hoboken, NJ: John Wiley & Sons, 2008:16–66.
17. Bewick V, Cheek L, Ball J. Statistics review 12: survival analysis. *Crit Care*. 2004;8:389–394.
18. Schober P, Bossers SM, Schwarte LA. Statistical significance versus clinical importance of observed effect sizes: what do P values and confidence intervals really represent? *Anesth Analg*. 2018;126:1068–1072.
19. Bland JM, Altman DG. The logrank test. *BMJ*. 2004;328:1073.
20. Hosmer DW, Lemeshow S, May S. Model development. In: *Applied Survival Analysis*. 2nd ed. Hoboken, NJ: John Wiley & Sons, 2008:132–168.
21. Bradburn MJ, Clark TG, Love SB, Altman DG. Survival analysis part III: multivariate data analysis—choosing a model and assessing its adequacy and fit. *Br J Cancer*. 2003;89:605–611.
22. Cox DR. Regression models and life-tables. *J R Stat Soc Series B*. 1972;34:187–220.
23. Hosmer DW, Lemeshow S, May S. Interpretation of a fitted proportional hazards regression model. In: *Applied Survival Analysis*. 2nd ed. Hoboken, NJ: John Wiley & Sons, 2008:92–131.
24. Nieto FJ, Coresh J. Adjusting survival curves for confounders: a review and a new method. *Am J Epidemiol*. 1996;143:1059–1068.
25. Hosmer DW, Lemeshow S, May S. Assessment of model adequacy. In: *Applied Survival Analysis*. 2nd ed. Hoboken, NJ: John Wiley & Sons, 2008:169–206.
26. Uno H, Claggett B, Tian L, et al. Moving beyond the hazard ratio in quantifying the between-group difference in survival analysis. *J Clin Oncol*. 2014;32:2380–2385.
27. Allison PD. Survival analysis. In: Hancock GR, Mueller RO, eds. *The Reviewer's Guide to Quantitative Methods in the Social Sciences*. New York, NY: Routledge, Taylor and Francis Group, 2010:413–424.
28. Fisher LD, Lin DY. Time-dependent covariates in the Cox proportional-hazards regression model. *Annu Rev Public Health*. 1999;20:145–157.
29. Huang WW, Zhu WZ, Mu DL, et al. Perioperative management may improve long-term survival in patients after lung cancer surgery: a retrospective cohort study. *Anesth Analg*. 2018;126:1666–1674.
30. Podolyak A, Sessler DI, Reiterer C, et al. Perioperative supplemental oxygen does not worsen long-term mortality of colorectal surgery patients. *Anesth Analg*. 2016;122:1907–1911.
31. Wilson SH, Wolf BJ, Bingham K, et al. Labor analgesia onset with dural puncture epidural versus traditional epidural using a 26-Gauge Whitacre needle and 0.125% bupivacaine bolus: a randomized clinical trial. *Anesth Analg*. 2018;126:545–551.
32. Kim M, Li G. Postoperative complications affecting survival after cardiac arrest in general surgery patients. *Anesth Analg*. 2018;126:858–864.
33. Chau A, Bibbo C, Huang CC, et al. Dural puncture epidural technique improves labor analgesia quality with fewer side effects compared with epidural and combined spinal epidural techniques: a randomized clinical trial. *Anesth Analg*. 2017;124:560–569.
34. Lee EK, Ahn HJ, Zo JI, Kim K, Jung DM, Park JH. Paravertebral block does not reduce cancer recurrence, but is related to higher overall survival in lung cancer surgery: a retrospective cohort study. *Anesth Analg*. 2017;125:1322–1328.
35. Windpassinger M, Plattner O, Gemeiner J, et al. Pharyngeal oxygen insufflation during AirTraq Laryngoscopy slows arterial desaturation in infants and small children. *Anesth Analg*. 2016;122:1153–1157.
36. Hosmer DW, Lemeshow S, May S. Parametric regression models. In: *Applied Survival Analysis*. 2nd ed. Hoboken, NJ: John Wiley & Sons, 2008:244–285.
37. Amorim LD, Cai J. Modelling recurrent events: a tutorial for analysis in epidemiology. *Int J Epidemiol*. 2015;44:324–333.
38. Dignam JJ, Zhang Q, Kocherginsky M. The use and interpretation of competing risks regression models. *Clin Cancer Res*. 2012;18:2301–2308.
39. Austin PC. A tutorial on multilevel survival analysis: methods, models and applications. *Int Stat Rev*. 2017;85:185–203.
40. Schober P, Vetter TR. Repeated measures designs and analysis of longitudinal data: if at first you do not succeed—try, try again. *Anesth Analg*. 2018;127:569–573.
41. Hosmer DW, Lemeshow S, May S. Other models and topics. In: *Applied Survival Analysis*. 2nd ed. Hoboken, NJ: John Wiley & Sons, 2008:286–354.
42. Schoenfeld DA. Sample-size formula for the proportional-hazards regression model. *Biometrics*. 1983;39:499–503.
43. Peduzzi P, Concato J, Feinstein AR, Holford TR. Importance of events per independent variable in proportional hazards regression analysis. II. Accuracy and precision of regression estimates. *J Clin Epidemiol*. 1995;48:1503–1510.