Graphing survival curve estimates for time-dependent covariates

LONNI R. SCHULTZ, EDWARD L. PETERSON, Department of Biostatistics and Research Epidemiology, Henry Ford Health Systems, Detroit MI, USA
NAOMI BRESLAU, Department of Psychiatry, Henry Ford Health Systems, Detroit MI, USA

ABSTRACT *Graphical representation of statistical results is often used to assist readers in the interpretation of the findings. This is especially true for survival analysis where there is an interest in explaining the patterns of survival over time for specific covariates. For fixed categorical covariates, such as a group membership indicator, Kaplan-Meier estimates (1958) can be used to display the curves. For time-dependent covariates this method may not be adequate. Simon and Makuch (1984) proposed a technique that evaluates the covariate status of the individuals remaining at risk at each event time. The method takes into account the change in an individual's covariate status over time. The survival computations are the same as the Kaplan-Meier method, in that the conditional survival estimates are the function of the ratio of the number of events to the number at risk at each event time. The difference between the two methods is that the individuals at risk within each level defined by the covariate is not fixed at time 0 in the Simon and Makuch method as it is with the Kaplan-Meier method. Examples of how the two methods can differ for time dependent covariates in Cox proportional hazards regression analysis are presented.*

Key words: Kaplan-Meier estimates, time dependent covariates, graphing procedures

Introduction

The use of survival analysis in psychiatric epidemiology is both widespread and well understood (for example, Anthony and Petronis, 1995; Kessler, Sonnega et al., 1995; Breslau, Johnson et al., 2001). It has been used to describe and compare the cumulative incidence of a psychiatric disorder or time to remission of a psychiatric disorder among distinct subpopulations using fixed covariates. It has also been used to assess the temporal association among different psychiatric disorders using time-dependent covariates. Our research group has performed survival analysis numerous times for both types of covariates (for example, Breslau, Schultz and Peterson, 1995; Breslau, Peterson et al., 1998; Breslau, Schultz et al., 2000).

Caution should be taken when choosing the appropriate method (fixed versus time-dependent covariates) to address the hypothesis of interest. A common mistake is to use fixed covariates when time-dependent covariates are more appropriate. Chilcoat and Breslau (1998) investigated this mistake for the relationship between post-traumatic stress disorder (PTSD) and drug-use disorder. Based on cross-sectional data and retrospective reports on age of onset of disorders, they compared the results of using fixed covariates to the results of using time-dependent covariates in Cox proportional hazards regression models. They identified the usefulness of using time-dependent covariates in the evaluation of the temporal relationship between PTSD and drug-use disorders by comparing the two sets of results from the survival analysis to prospective results in the same population. They also demonstrated how the use of fixed covariates in the models can give misleading results.

As in other methods of analysis, it is important both to tabulate the data and to provide graphical representations of the survival analysis results. For a fixed categorical variable, the product limit method of Kaplan and Meier (1958) produces a survival curve for each group, defined by the levels of the covariate. However, for the case of a time-dependent categorical covariate the representation in the literature is not so

clear. In many cases, no graphical representation is presented for time-dependent covariates, limiting the ability of the author to convey the results clearly. In other cases, a transformation is made representing the time-dependent covariate as a fixed covariate. This is often done by assigning the lifetime status of the time-dependent covariate and treating it as fixed at time 0. Kaplan-Meier survival curves are then generated from this transformed variable. This graphical approach can dramatically misrepresent the data and obscure the results.

Simon and Makuch (1984) suggest a computational method that can be used to graphically represent survival curves for time-dependent covariates. This method appears to be similar to the one used by Stata (2001) to compute 'Kaplan-Meier' estimates for time-dependent covariates, where multiple records per individual are used to address the temporal sequence of events.

In this paper we review the product-limit method of Kaplan and Meier for producing survival curves for fixed covariates. We contrast this to the methods proposed by Simon and Makuch for graphing time dependent covariates, using Kaplan-Meier type calculations. Finally, we present two examples that illustrate the usefulness of the time-dependent graphing approach.

Methods

For n subjects, let $t_1 < t_2 < \ldots < t_k$ ($k \leq n$) denote the distinct observed event times. Let n_i be the number of subjects at risk at a time just prior to t_i and d_i be the number of events at t_i. The Kaplan-Meier estimate for the survivor function S(t) is a discrete function estimated by

$$\hat{S}(t) = \prod_{t_i < t} \frac{n_i - d_i}{n_i}$$

A separate Kaplan-Meier curve would be computed for each level of the covariate being considered. Each individual subject is assigned at time 0 to a unique group by his or her covariate value (for example 'exposed to risk factor' versus 'not exposed'). Hence, the number of individuals at risk for the event decreases over time due to the individual experiencing the event or being lost to follow up. An individual subject remains in the same group to which he/she was assigned at time 0 throughout the analysis.

Simon and Makuch propose a method to generate the survival curves for the different levels of a time-dependent covariate. The estimation of the survival curve uses the same formula as the Kaplan-Meier approach. The difference imposed by considering a time-dependent covariate is in the interpretation of the risk set. A fixed covariate has uniquely defined risk sets at the start of the study, which do not overlap. In contrast, a time-dependent covariate has changing risk sets that are redefined, based on the current value of the covariate, at each unique event time. Hence, the risk set n_i consists of all individuals who are at risk just prior to time t_i and whose covariate value indicates membership in the group being considered at time t_i. This number can increase or decrease over the time course covered in the study.

A characteristic of the Simon and Makuch method should be mentioned here. The number of subjects at risk is usually small in at least one level of the covariate in the early time points, so Simon and Makuch suggest selecting a starting time point after the study's time 0 for graphical presentation. This time point would be selected to assure an adequate number of subjects at risk in each level of the covariate and would depend upon the distribution of events. Events that occur before the chosen starting time point are ignored.

Examples

Example 1. In a sample of young adult smokers (n = 424), we assessed the effect of alcohol-use disorder on smoking cessation (Breslau, Peterson et al., 1996). The lifetime history of alcohol use disorder among smokers could have taken three forms:

- never had an alcohol use disorder;
- have an alcohol use disorder; and
- had an alcohol use disorder in the past but had been remitted.

Each subject's status could change over time. For example, a subject started out with no alcohol-use disorder, developed an alcohol use disorder at age 20 and recovered at age 25. This subject has been in all three alcohol-disorder groups, depending on the age being considered. Hence, the alcohol-use disorder status is a time-dependent covariate.

The starting point for all subjects was their age of onset of daily smoking. The endpoint of interest was time of smoking cessation. Subjects were censored if

they had not stopped smoking for at least one year prior to follow-up. Number of years from daily smoking until smoking cessation or last follow-up was the time variable.

Cox regression with time-dependent covariates was applied to test the differences among the three alcohol groups with respect to the likelihood of quitting smoking. An overall difference among the three groups was detected ($p = 0.0012$, Wald chi-square test). Table 1 contains the results of the pairwise comparisons among the three groups. The subjects who remitted from an alcohol-use disorder were most likely to stop smoking, whereas the subjects who had an active alcohol use disorder and did not remit were least likely to stop smoking.

Cox regression analysis was also performed in which alcohol status was defined as fixed according to an individual's alcohol status at the end of follow-up. Using fixed covariates, an overall difference among the three alcohol groups was not significant ($p = 0.3707$, Wald chi-square test). The results of the pairwise comparisons among the three groups for fixed covariates are given in Table 1.

Figures 1 and 2 graphically represent the survival curves for the three groups using the time dependent and fixed covariates methods. Figure 1 displays the three curves generated for the time-dependent covariate, using the method of Simon and Makuch. In this figure, a subject's alcohol status was determined each time someone stopped smoking. Figure 1 reflects accurately the findings from the Cox regression with time dependent covariates presented in Table 1. It shows that the subjects whose alcohol-use disorders has remitted are more likely to stop smoking, followed by the subjects who never had an alcohol use disorder, with the subjects who still have an active alcohol use disorder being least likely to stop.

Figure 2 contains the fixed covariate curves for the three groups. Each subject's alcohol disorder status was defined as either never having an alcohol-use disorder, having an alcohol-use disorder, and remitted alcohol use disorder. This status was set at the time of onset of daily smoking and did not change over time. The figure shows no differences among the three groups.

Example 2. Using the results of Chilcoat and Breslau, the association between drug-use disorder and PTSD was explored. The following analyses assessed the effect of drug-use disorder on PTSD, which corresponds to the results presented in Table 5 of their paper. The hazard ratios and confidence intervals will differ slightly from Chilcoat and Breslau's results because in this example there were no adjustments for sex, race and education.

In this analysis, the subjects were divided into two groups depending on their drug-use disorder ('present' versus 'absent'). As in example 1, the variable was time dependent, the onset of drug use disorder preceding the time of trauma and PTSD in some cases but not all. The starting point (time 0) was birth, with the end point of interest being age at PTSD. Subjects who had not experienced PTSD at the end of follow-up

Table 1. Pairwise comparisons of alcohol-use disorder status on smoking cessation from Cox regressions

	Time dependent covariates		Fixed covariates	
Comparison of interest	Hazard ratio	95% CI	Hazard ratio	95% CI
Continue versus never	0.44	(0.26, 0.76)	0.72	(0.43, 1.20)
Remitted versus never	1.54	(0.94, 2.52)	0.84	(0.57, 1.25)
Remitted versus continue	3.45	(1.58, 7.54)	1.18	(0.66, 2.11)

Table 2. Comparisons of drug-use disorder status (present versus absent) on PTSD from Cox regression

	Time dependent covariates		Fixed covariates	
Comparison of interest	Hazard ratio	95% CI	Hazard ratio	95% CI
Present versus absent	1.51	(0.80, 2.87)	2.33	(1.52, 3.56)

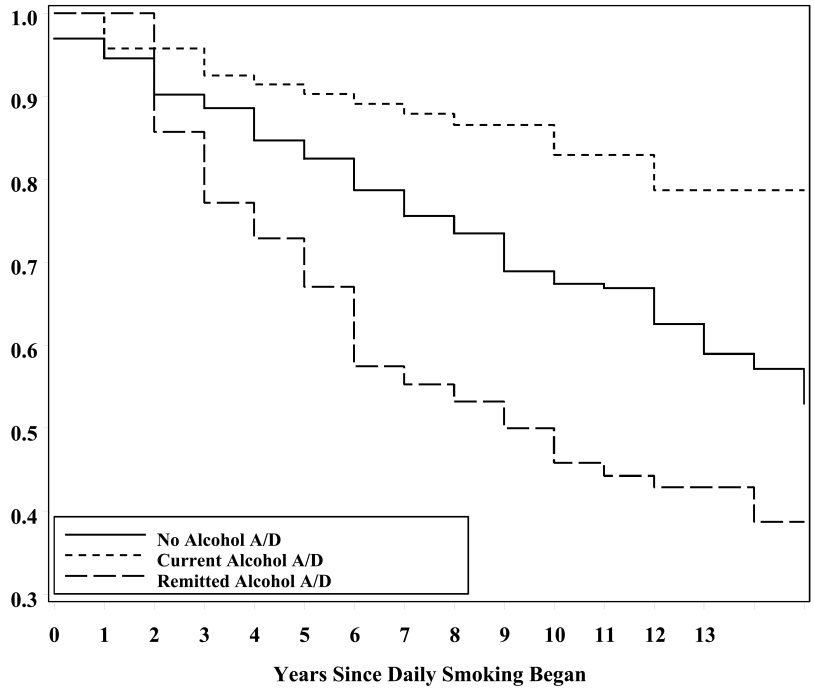

Figure 1: Survival curves of time to smoking cessation for alcohol use disorder status (time dependent covariate) using the Simon and Makuch method.

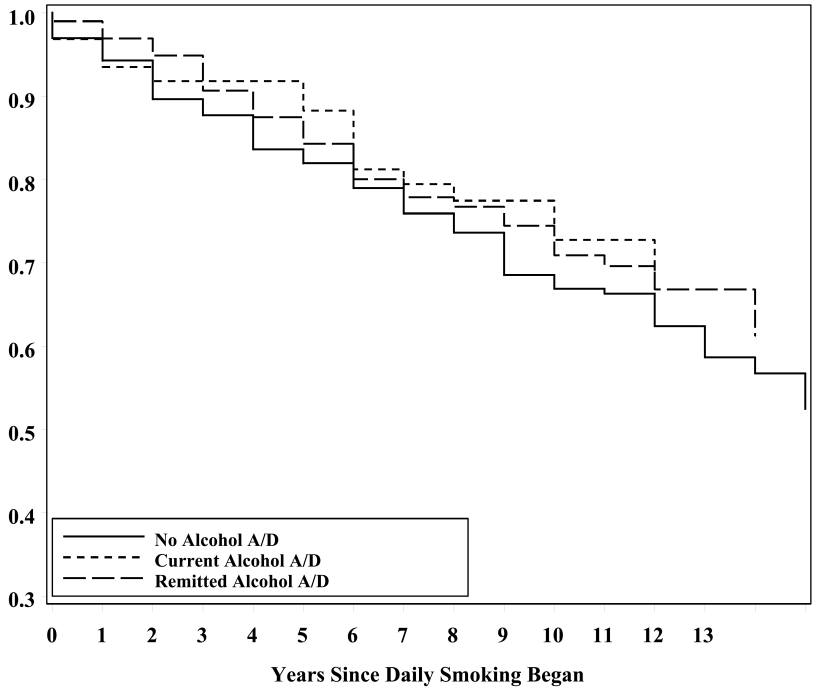

Figure 2. Survival curves of time to smoking cessation for alcohol use disorder status (fixed covariate) using the Kaplan-Meier method.

were censored at the age they were at that time. Chronological age was the time variable. Two survival approaches were used, with and without time-dependent covariates.

Cox regression with a time-dependent covariate for drug use disorder was done to assess the effect of the presence of drug use disorder on PTSD. The effect of drug use disorder was not statistically significant ($p = 0.204$, see Table 2).

Cox regression was also estimated using a fixed covariate for drug use disorder. If an individual ever experienced a drug-use disorder during their lifetime, their status was set to 'drug-use disorder present'. The value of this status was used from time 0 (birth) forward. Using a fixed covariate analysis, the effect of drug-use disorder was statistically significant ($p < 0.0001$, also see Table 2).

Figures 3 and 4 graphically represent the survival curves for drug use disorder using the time dependent and fixed covariate methods, respectively. As in example 1, these figures correspond to the findings from the Cox regression analysis for their corresponding type of covariate.

Discussion

In psychiatric epidemiology, a number of the diagnostic tools (such as DIS4, or CIDI) are used to collect lifetime data on psychiatric disorders. From the data, we are interested in assessing the temporal order of two or more of these disorders. Cox proportional hazards regression with time-dependent covariates can be used to quantify the effects of one disorder on the likelihood of another one, while taking into account the sequence of the disorders. We also desire to graph the results to present more clearly the patterns of the findings across time.

From the above examples, we see that the use of fixed covariates estimates when the covariate is time dependent could produce misleading information regarding the relationship between the covariate and the event of interest. Unlike the Kaplan-Meier method, the Simon and Makuch method does not make the assumption that the independent covariate status is set and fixed at time 0. The graphs from this method also agree with the findings of the Cox proportional hazard regression with time dependent covariates.

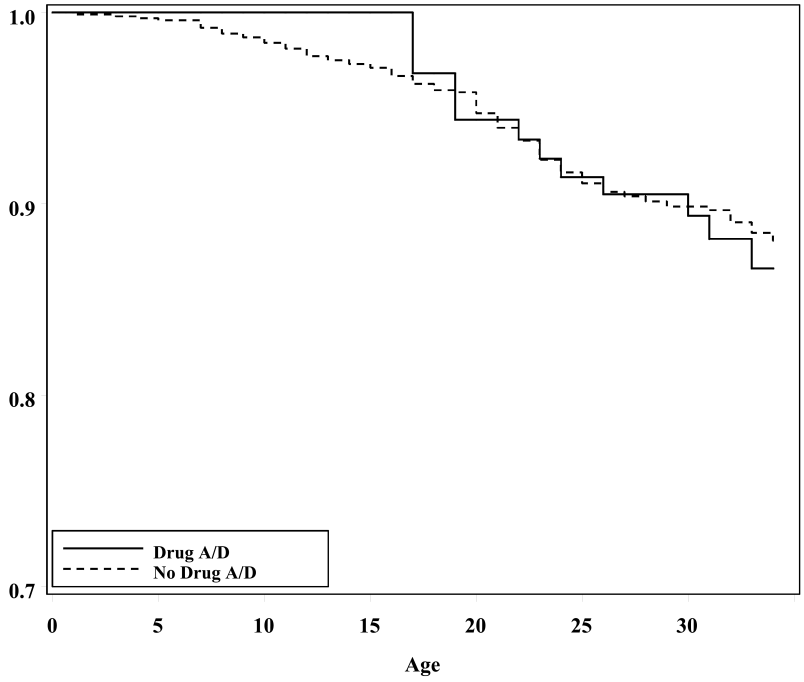

Figure 3. Survival curves of time to PTSD for drug use disorder status (time dependent covariate) using the Simon and Makuch method.

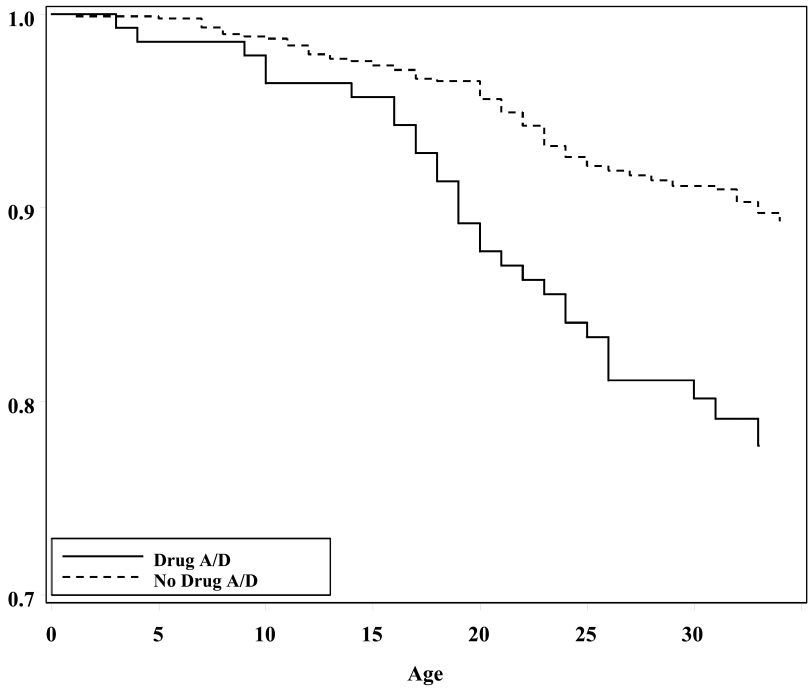

Figure 4. Survival curves of time to PTSD for drug use disorder status (fixed covariate) using the Kaplan-Meier method.

In both examples, the value of the time-dependent covariate at time t was determined to be the status of the covariate the year prior to time t. In other analyses, different covariate values could be defined to incorporate specific covariate characteristics of interest, such as delayed effects or cumulative effects.

In the examples presented here, the effect of a single covariate was assessed. In most psychiatric research analyses, adjustments for other covariates are usually necessary or desired. The Simon and Makuch method, like the Kaplan-Meier method, can be computed for different subsets of the study population in a stratified manner. Neither method allows for the direct adjustment of other covariates when assessing a single covariate. Stata does give survival estimates that are adjusted for other covariates. However, these estimates are based on the baseline survivor function from the Cox proportional hazards regression models performed within each level of the covariate of interest.

A limitation of the Simon and Makuch method is the determination of a starting point for the graphical presentation. The choice of the starting point is data dependent and may change from study to study, even if the independent covariates and event of interest do not change. Hence the starting point should be selected carefully. In the examples presented above, the starting points were not specified and the graphs started at time 0. For example 2, it appears that a starting time after age 15 may be more appropriate than time 0, as very few individuals experience drug-use disorders before this age.

Another limitation of the Simon and Makuch method is the survival estimate could be zero at a given time and will remain zero for the remaining time points, regardless of the fact that other subjects may be at risk at later times. Survival estimates of zero occur when all of the subjects at risk for a given covariate status experience the event at that time. The potential for zero survival is reduced with the selection of an appropriate starting point to insure an adequate number of subjects at risk.

Even with these limitations, the Simon and Makuch method produces a more accurate graphical representation of the difference in survival curves then assuming a fixed covariate and plotting the Kaplan-Meier curves. It appears to be very useful for graphically representing the quantified effects from the Cox regression with time dependent covariates.

Acknowledgements

This work was supported by grant P50 NS32399 of the United States National Institute of Health Headache Research Center, Bethesda, Maryland (KMA Welch, MD).

References

Anthony JC, Petronis KR. Early-onset drug use and risk of later drug problems. Drug and Alcohol Dependence 1995; 40: 9–15.

Breslau N, Schultz L, Peterson E. Sex differences in depression: a role for pre-existing anxiety. Psychiatry Research 1995; 58: 1–12.

Breslau N, Peterson EL, Schultz LR, Andreski P, Chilcoat H. Are smokers with alcohol disorders less likely to quit? American Journal of Public Health 1996; 86: 985–90.

Breslau N, Peterson EL, Schultz LR, Chilcoat H, Andreski P. Major depression and stages of smoking. Archives of General Psychiatry 1998; 55: 161–6.

Breslau N, Schultz LR, Stewart WF, Lipton RB, Lucia VC, Welch KMA. Headache and major depression: is the association specific to migraine? Neurology 2000; 54: 308–13.

Breslau N, Johnson EO, Hiripi E, Kessler R. Nicotine dependence in the United States: prevalence, trends and smoking persistence. Archives of General Psychiatry 2001; 58: 810–16.

Chilcoat H, Breslau N. Investigations of causal pathways between PTSD and drug use disorders. Addictive Behaviors 1998; 23: 827–40.

Kaplan EL, Meier P. Nonparametric estimation from incomplete observations. Journal of American Statistical Association 1958; 53: 457–81.

Kessler RC, Sonnega A, Bromet E, Hughes M, Nelson CB. Posttraumatic stress disorder in the National Comorbidity Survey. Archives General Psychiatry 1995; 52: 1048–60.

Simon R, Makuch RW. A non-parametric graphical representation of the relationship between survival and the occurrence of an event: application to responder versus non-responder bias. Statistics in Medicine 1984; 3: 35–44.

StataCorp 2001. Stata Statistical Software: Release 7.0 College Station, TX: Stata Corporation.

Correspondence: Lonni Schultz, Department of Biostatistics and Research Epidemiology, Henry Ford Health Systems, One Ford Place, 3E Detroit MI 48202, USA.
Telephone: +1 (313) 874-6364.
Fax: +1 (313) 874-6730.
Email: lschult1@hfhs.org.